The Use of Mixture Models in Capture–Recapture

Author:
Chen Yu (郁　辰)

Supervisors:
Prof. Byron Morgan
Dr. Diana Cole

A thesis submitted for the degree of

Doctor of Philosophy

in the subject of

Statistics

School of Mathematics, Statistics and Actuarial Science

August 2015

Abstract

Mixture models have been widely used to model heterogeneity. In this thesis, we focus on the use of mixture models in capture–recapture, for both closed populations and open populations. We provide both practical and theoretical investigations. A new model is proposed for closed populations and the practical difficulties of model fitting for mixture models are demonstrated for open populations. As the number of model parameters can increase with the number of mixture components, whether we can estimate all of the parameters using the method of maximum likelihood is an important issue. We explore this using formal methods and develop general rules to ensure that all parameters are estimable.

Acknowledgments

I'd like to thank my parents, Mr Lianhua Yu (郁联华) and Miss Qiying Zhou (周琦瑛), who have supported me all the way here. It is not only the financial support, but also for having my back and a home that I can always come back to.

I'd like to express my gratitude to my academic supervisors Professor Byron J.T. Morgan and Dr Diana J. Cole. I thank Byron for reading my thesis again and again, correcting my spelling, the use of propositions, and academic writing. I thank Diana for helping me developing the main part of the research as well as the use of Maple. I'd like to thank Professor Shirley Pledger for providing the open population data and Dr Ruth King for providing closed population data.

I'd also like to thank the National Centre for Statistical Ecology (NCSE) and the School of Mathematics, Statistics and Actuarial Science (SMSAS) for jointly funding my PhD, without which, it would not be possible.

I'd like to thank all the users in the online discussion forums `MaplePrime` and the Mathematica® and LaTeX forum on `stackexchange`. They have been really helpful. I'd like to thank the Chinese Society, the Kent Graduate Student Association (KGSA), KentUnion and the International development office, for keeping me away from only study. So that life is not just about sitting in front of a computer from 9am to 5pm.

Last but not least, I'd like to thank all the administrative staff as well as teaching staff in the School of Mathematics, Statistics and Actuarial Science (SMSAS), for everything! Thank you for having me and it has been great experience!

Contents

Abstract i

Acknowledgments ii

1 Background and Introduction 1

1.1 Statistical Ecology & Capture–Recapture . . . 1

1.2 Relevant History of Capture–Recapture Models and Parameter Redundancy . . . 2

1.3 Structure of Thesis . . . 5

1.4 Use of Computer Packages . . . 8

2 Mixture Models in Closed Populations 11

2.1 What is a Closed Population? . . . 11

2.2 Data and Models . . . 12

2.3 Model Fitting . . . 18

2.4 Results . . . 22

2.5 Conclusion . . . 25

2.6 Code & List of files . . . 26

3 Mixture Models in Open Populations 27

3.1 Introduction . . . 27

3.2 capture–recapture . . . 28

3.2.1 Background . . . 28

3.2.2 Data . . . 28

3.2.3 Cormack–Jolly–Seber Model . . . 30

3.2.4 Jolly–Seber Model . . . 37

3.3 Capture–Recovery . . . 40

3.3.1 Background . . . 40

3.3.2 Data . . . 42

3.3.3 Model . . . 42

3.4 Capture–Recapture–Recovery . . . 43

3.4.1 Background . . . 43
3.4.2 Data . . . 44
3.4.3 Model . . . 46
3.5 Code & List of files . . . 47

4 Parameter–Redundancy in Open Mixture Models For capture–recapture 50
4.1 What is parameter–redundancy? . . . 51
4.2 Methods for Detecting parameter–redundancy . . . 53
4.2.1 Background . . . 53
4.2.2 Notation . . . 53
4.2.3 Symbolic Method . . . 54
4.2.4 Hybrid Symbolic–Numerical (SN) Method . . . 62
4.3 Results Using the SN Method . . . 68
4.4 Theorems Needed for Formal Proof . . . 75
4.5 Conclusion . . . 78
4.6 Code & List of files . . . 79

5 Simpler Exhaustive Summaries 82
5.1 Using the Linear Link Function . . . 84
5.1.1 How to find a reparameterisation? . . . 84
5.1.2 Model: $\{\phi(t), p(t + h_C)\}$ . . . 92
5.1.3 Sub–Models of the model $\{\phi(t), p(t + h_C)\}$ . . . 103
5.1.4 Model: $\{[\phi(h), p(h)]_C\}$ . . . 106
5.1.5 Model: $\{[\phi(h), p(t + h)]_C\}$ . . . 110
5.1.6 Model: $\{\phi(h_C), \lambda(t)\}$ . . . 112
5.1.7 Model: $\{\phi(h_C + a), \lambda(t)\}$ (with covariates) . . . 114
5.1.8 Introducing the $H(x, y, z)$ function . . . 117
5.1.9 Model: $\{\phi(h_C), p(h_C), \lambda(h_C)\}$ . . . 118
5.2 Logistic Link . . . 121
5.2.1 One–to–one transformations between the linear link function and the logistic link function . . . 121
5.2.2 Model: $\{\phi(\cdot), p(t + h_C)\}$ . . . 124
5.2.3 Model: $\{\phi(t), p(t + h_C)\}$ . . . 127
5.3 Relationship to Parameter Redundancy . . . 131
5.3.1 Tables of Simpler Exhaustive Summaries . . . 131
5.3.2 Tables of Parameter Redundancy Results . . . 133
5.4 Conclusion . . . 137
5.5 List of files . . . 138

6 Fitting Mixture Models to Open Population Data **139**
6.1 Fitting Real Datasets . . . 141
6.1.1 Brushtail Possum (*Trichosurus vulpecula*) Data . . . 141
6.1.2 Optimization Routines . . . 141
6.1.3 Example: $\{\phi(t), p(t+h_2)\}$ . . . 143
6.1.4 Example: $\{[\phi(t+h), p(t+h)]_2\}$ (Similar & Opposite) . . . 149
6.1.5 Example: $\{\phi(t+h_2), p(t+h_2)\}$ Four–class cross–classified model . . . 152
6.2 Model Selection . . . 156
6.3 Using Different Link Functions . . . 161
6.4 Models With Interactions . . . 168
6.5 Conclusion . . . 171
6.6 List of files . . . 173

7 Conclusion and Future Work **175**
7.1 Conclusion . . . 175
7.2 Future Work . . . 177

A Maple® Computer Package: `MixtureModelforCaptureRecapture` **178**
A.1 About the package . . . 178
A.2 How to use the package? . . . 178
A.3 Functions Available . . . 179
A.4 The `mykappa` function . . . 180
A.5 Using a real dataset . . . 181

B Mathematica® Computer Package: `MixtureModelforCaptureRecapture` **184**
B.1 About the package . . . 184
B.2 How to use the package? . . . 184
B.3 Functions Available . . . 185
B.4 Usage . . . 185

List of Figures

2.1 An illustration of the setup of the closed population model . . . 12
2.2 A comparison of model fits for the `link3` dataset. Here we show how the four models describe the capture probability near zero. 25

3.1 Illustration of capture histories for CJS models of capture–recapture data . . . 33

4.1 The derivative matrix from Gimenez *et al.* (2003), for which Maple® would give an incorrect answer . . . 61
4.2 An example of Maple® giving the wrong rank. . . . 63
4.3 An example of Maple® giving the correct rank . . . 64
4.4 An example of Mathematica® giving the correct rank . . . 65
4.5 A general framework for detecting parameter–redundancy using exhaustive summary . . . 76

6.1 Profile plots of $\phi_j, j = 1, \ldots, 8$ for model $\{\phi(t), p(t + h_2)\}$ . . . 147
6.2 Profile plots of μ_p, η_p, w and $\tau_j, j = 3, \ldots, 9$ for model $\{\phi(t), p(t+h_2)\}$ . . . 148
6.3 Illustration of the "similar" model and the "opposite" model. . . 151
6.4 Profile plot of $\eta_{\phi 2}$ for model $\{[\phi(t + h), p(t + h)]_2\}$ . . . 155
6.5 Illustration of the four–class cross–classified model $\{\phi(t+h_2), p(t+h_2)\}$ . . . 155

A.1 How to install the Maple® package? . . . 179

List of Tables

2.1 Examples of closed population data sets 13
2.2 North American Breeding Bird Survey (BBS) data for year 1997, 1998, 1999, 2000 and 2001 14
2.3 Summary of closed population models 18
2.4 Rules of Gaussian Quadrature (Stroud and Secrest, 1966) 19
2.5 Maximized log likelihood for closed population datasets 22
2.6 AIC values for closed population data sets 24
2.7 List of files on Chapter 2 26

3.1 Capture–recapture data in an m–array form 29
3.2 Numbers of model parameters with K samples and C classes for CJS mixture models 38
3.3 Numbers of model parameters with K samples and C classes for JS mixture models 41
3.4 List of files on Chapter 3 47

4.1 The dimensions of the derivative matrices 59
4.2 Model ranks obtained from using the SN method for model $\{[\phi(h), p(h)]_C\}$ for different dimensions of K and C shown. . . . 70
4.3 Model ranks obtained from using the SN method for model $\{\phi(t), p(t \times h_C)\}$ for different dimensions of K and C shown. . . 71
4.4 Different model ranks obtained from using the linear link function and the logistic link function using the SN method for K and C shown. 74
4.5 List of files on Chapter 4 80

5.1 Model ranks obtained from using the SN method for model $\{\phi(t), p(t + h_C)\}$ for different dimensions of K and C shown. . . 93
5.2 Summary of estimable parameters in terms of $\boldsymbol{s}(\boldsymbol{\theta})$ for $K = 3, 4, 5, 6$ for the model $\{[\phi(h), p(t + h)]_C\}$ for any $C \geq 1$. 111
5.3 Summary for $K = 2, \ldots, 5$ for model $\{\phi(h_C), \lambda(t)\}$ for any $C \geq 1$. 113

5.4 The elements of $\boldsymbol{\kappa}(\boldsymbol{s})$ for $K = 2, \ldots, 5$, for model $\{\phi(h_C + a), \lambda(t)\}$ for any $C \geq 1$. 116
5.5 Estimable parameter combinations and simpler exhaustive summaries for $K = 2, \ldots, 5$ for model $\{\phi(h_C + a), \lambda(t)\}$ for any $C \geq 1$. 117
5.6 Terms that cannot be estimated separately from the model $\{\phi(h_C), p(h_C), \lambda(h_C)\}$ for $K = 2, \ldots, 6$ 119
5.7 Terms that cannot be estimated separately from the model $\{\phi(h_C), p(h_C), \lambda(h_C)\}$ for $K = 2, \ldots, 6$ (written in terms of $H(x, y, z)$ 120
5.8 Simpler exhaustive summaries (CJS mixture, linear link): heterogeneity exists only in ϕ or p 132
5.9 Parameter redundancy results (CJS mixture models with linear link function): heterogeneity exists only in ϕ or p 135
5.10 Parameter redundancy results (JS mixture models with linear link function): heterogeneity exists only in ϕ or p 136
5.11 List of files on Chapter 5 138

6.1 Estimated parameters for model $\{\phi(t), p(t + h_2)\}$ 146
6.2 Point estimates for $\{\phi(t), p(t + h_2)\}$ 150
6.3 Maximum–likelihood estimates for $\{[\phi(t + h), p(t + h)]_2\}$ 153
6.4 Point estimates for $\{[\phi(t + h), p(t + h)]_2\}$ 154
6.5 Comparison of results for model $\{\phi(t + h_2), p(t + h_2)\}$ 157
6.6 Summary of model fitting of female possum data using the logistic link function 158
6.7 Summary of model fitting of male possum data using the logistic link function 159
6.8 Summary of model fitting of combined possum data using the logistic link function 160
6.9 Comparison of model AIC values from using logistic link and linear link 162
6.10 List of files on Chapter 6 174

A.1 Functions in the Maple® package `MixtureModelforCaptureRecapture.mla` 180
A.2 Optional arguments for Maple® function `mykappa` 182

B.1 Functions in the Mathematica® package `MixtureModelforCaptureRecapture.m` 186
B.2 Optional arguments for Mathematica® function `mykappa` 187

Chapter 1

Background and Introduction

1.1 Statistical Ecology & Capture–Recapture

In this thesis we focus on statistical models in ecology. We are interested in knowing the population size of species of interest, the survival probabilities of individuals, and so on. Ecologists take regular samples, which are often annual, to collect data from populations. The sampling method of interest to us is called *capture–recapture*.

In capture–recapture, individuals are marked, released, and recaptured, or recovered, typically at equal time intervals. Usually each individual has a unique tag for identification. After initial capture, if an individual is seen again alive, this is a *recapture*. If an individual is found again dead, this is a *recovery*. By taking repeated samples, we build up a capture history for each individual in the population, except for those we never observe again of following initial capture and marking.

1.2 Relevant History of Capture–Recapture Models and Parameter Redundancy

First of all, consider a closed population, where no birth, no death and no migration occur over the sampling period. The Lincoln–Petersen (Lincoln, 1930; Petersen, 1896) estimate, which assumes a closed population, is obtained when there are just two samples, and estimates the population size N using the equation below

$$\frac{n_1}{N} = \frac{m_2}{n_2},$$

in which n_1 denotes the size of the first sample all of which were marked, and n_2 denotes the size of the second sample, of which m_2 were marked. The closed population means we assume the totol population does not change between the two visits when the samples are taken.

By taking repeated samples of a closed population, the data we collected form what is called a *Schnabel census* (Sutherland, 1996). Assuming an equal capture probability for all individuals on each occasion give us the simplest M_0. When we have the capture histories for each individual, we can allow individuals to have different capture probabilities and we denote the resulting heterogeneity model M_h. If we allow temporal variation of capture probabilities, then we denote the model as M_t. If we allow individuals to have a different capture probability after initial capture, we have a behaviour–related model, denoted by M_b. This might be appropriate if animals are trap–happy or trap–shy, following first capture.

We can also allow both time variation and heterogeneity in the model denoted by M_{th}. Many other closed population models can also be considered, see Otis *et al.* (1978). For example, we can introduce both the time effect and the behavioural effect, denoted by M_{tb}. The full model, M_{tbh}, allows for time,

behavioural and heterogeneous effects on the capture probability. If we allow individuals to have different capture–recapture probabilities and/or survival probabilities, we can have mixture models. Mixture versions of such models are considered in Pledger (2000). Pledger (2000) also considers models with interactions between the effects, with the full model denoted by $M_{t\times b\times h}$. This model has time, behavioural and heterogeneous effects, as well as interactions between time and behavioural, time and heterogeneity, and behavioural and heterogeneity effects. When a model contains mixture components, we have to consider whether all model parameters are estimable: the issue has been discussed in Link (2003) and Holzmann *et al.* (2006).

When the individual capture histories are not accessible to us, instead we only have the numbers of animals that were captured j times, denoted by f_j, for $j = 1, 2, \ldots, K$, where K is the length of the capture–recapture study. We can assume that all individuals have the same recapture probabilities, so that for each f_j, p_j describes the probability of f_j. This will allow us to form a multinomial distribution on $\{f_j\}$ and we can model heterogeneity of recapture probabilities through the use of different distributions. For example, we can model $\{p_j\}$ using a logistic normal binomial distribution or a beta–binomial distribution. Further, we can consider mixtures of distributions, see Morgan and Ridout (2008). For example, a mixture of a binomial distribution and a beta–binomial distribution has been used.

In contrast, in open populations we allow individuals to enter the population through birth and leave the population through death and in addition there may be migration. Again, our focus is to model heterogeneity if it is present through the repeated samples collected. We allow individuals to vary from each other, by assuming there are a number of groups each with constant parameters but with variation between groups. Mathematically, this is

modelled by having mixture components. We denote the length of the capture–recapture study by K, and the number of mixture components by C.

Pledger (2000) discussed the use of mixture models in closed populations and noted that it might not be possible to estimate all model parameters given the data collected. Holzmann *et al.* (2006) have done a further study to prove that a necessary condition to be able to estimate all model parameters is $K \geq 2C$ for the types of closed population mixture models in Pledger (2000). Although open population mixture models have been used widely (Pledger *et al.*, 2010, 2003; Pledger and Schwarz, 2002), whether we can estimate all model parameters in these mixture models has not been formally studied. The general rule to ensure that models are identifiable and that all parameters in the closed–population are estimable was derived in Holzmann *et al.* (2006), but no previous work has formally explored the rules of mixture models in open–populations. It was noted by the authors in Pledger *et al.* (2010) that

> Finding exact rules for these more complex open–population models would be a difficult and time–consuming exercise, and the rules are unlikely to be simple.

This motivates the main part of this thesis in Chapter 4 and Chapter 5. Here we will be focusing on finding rules for open–population capture–recapture mixture models.

A formal definition of model *identification* can be found in Silvey (1975). The problem of whether we are able to estimate all model parameters is known as *parameter redundancy*. There are two types. If the problem occurs due to the inherited model structure, this is called *intrinsic parameter redundancy*, while if the problem occurs due to a particular dataset, this is called *extrinsic parameter redundancy*; see Hubbard (2014). In this thesis, we only consider the

former case, and investigate parameter redundancy resulting from the model structure.

Methods of using symbolic algebra for detecting parameter redundancy include Catchpole *et al.* (1998, 2002) and Catchpole and Morgan (1997, 2001). The quantity that uniquely determines the model structure is a set of parameters called an *exhaustive summary*. We focus on methods that are based on *exhaustive summaries*, see Cole *et al.* (2010), which has developed the theory based on previous work. These methods have been used for many non–mixture models. For example, see Hubbard (2014); Cole *et al.* (2012); Hubbard *et al.* (2014) for mark–recovery models, Cole and Morgan (2010a) for tag return models, and Cole (2012) for multi–state capture–recapture models.

When the models are too complex for the symbolic method we shall employ, it is also possible to use a numerical approach to detect parameter redundancy (Viallefont *et al.*, 1998). For example see Jiang *et al.* (2007) for tag return models and Hunter *et al.* (2009) for multi–state mark–recapture models. We can also use a hybrid symbolic–numerical (SN) method; see Choquet and Cole (2012).

Not only do we discuss theoretical parameter redundancy results, but we also discuss the practical model fitting of these open–population capture–recapture mixture models. For capture–recapture, such models have been fitted in Pledger *et al.* (2010, 2003) and for capture–recovery, the models have been fitted in Pledger and Schwarz (2002). We compare our results to published results and by determining how many mixture components we can have, we also fit more plausible models. Sometimes we are able to find a simpler model.

1.3 Structure of Thesis

The main content of this thesis consists of five chapters, from Chapter 2 to Chapter 6. We give brief introductions to each of these chapters below.

In Chapter 2, we start by looking at the performance of a variety of mixture models in the closed–population scenario. We give the assumptions of the closed–population in a statistical ecology environment and explain how the data are collected. By setting up a multinomial model we can estimate the size of the population.

The probabilities in the multinomial model are modelled through different mixtures of probability distributions. Link (2003) discussed the use of the logistic normal binomial model as well as the log–gamma model. Morgan and Ridout (2008) explore the behaviour of certain mixture binomial models: the beta binomial model, and a mixture of a binomial and a beta binomial model. We introduce a new mixture, the binomial–logistic normal binomial model. The performance of this new model is compared with the previous models given above.

The remaining part of the thesis focuses on parameter–redundancy in open capture–recapture models. Chapter 3 introduces the models and gives examples of the numbers of parameters considered in such models. Chapter 4 gives a definition of parameter–redundancy and describes the methods that we use to detect the parameter–redundancy, and how we proceed when it is present, based on the use of exhaustive summaries. Chapter 5 gives examples of how we can simplify the problem by introducing what are known as simpler exhaustive summaries. Chapter 6 demonstrates the practical difficulty of fitting open capture–recapture mixture models.

Chapter 3 lists all the open capture–recapture mixture models we consider in this thesis. Models of interest are capture–recapture models (Pledger *et al.*, 2003, 2010), capture recovery models (also known as tag–return, band–recovery or mark recovery) (Pledger and Schwarz, 2002) and capture–recapture recovery models. For each model, we give a gentle introduction, followed by an illustration of the data being described. We then give the expressions of the different mixture models.

Chapter 4 discusses the problem of parameter–redundancy in open capture–recapture mixture models. We list the tools for detecting the problem. Methods include symbolic approaches developed in Catchpole and Morgan (1997); Catchpole *et al.* (1998) and Cole *et al.* (2010), and the symbolic–numerical hybrid method in Choquet and Cole (2012). For the first time, results of parameter–redundancy are given for small values of K and C for some mixture models as examples, where as mentioned earlier, K is the length of the capture–recapture study and C is the number of mixture components. We present theorems and a general framework to approach the problem of parameter–redundancy based on the use of exhaustive summaries.

Chapter 5 consists of various examples of what are known as simpler exhaustive summaries for open capture–recapture mixture models. We provide examples using both the linear link function and the logistic link function. We discuss the use of simpler exhaustive summaries and how we can use them to simplify the problem of detecting and proving parameter–redundancy results. The chapter involves heavy algebraic manipulations and some mathematical expressions have been omitted. We use symbolic computer packages to do most calculations and all relevant code is included in the supplementary CD. Brief guides on the packages written for the thesis are given in Appendix A for the computer codes in Maple® and Appendix B for the computer codes in

Mathematica® .

Chapter 6 shows the fitted results for open capture–recapture mixture models. We use examples to demonstrate the difficulty of fitting these models. We compare our fitted results to the results in Pledger *et al.* (2003). We further discuss how to select the best model(s). We also compare the fitted results from using different link functions, such as the linear link function and the logistic link function. Models with interactions are also discussed. We find better and simpler models than those published.

Chapter 7 summarises the thesis and discusses further work that needs to be done.

1.4 Use of Computer Packages

Statistics is just like any other science: the research typically relies on using a computer. In this thesis, complex mixture models are considered, where most of the models are not readily available in any of the current computer packages, so coding cannot be avoided.

Throughout the thesis, we provide computer codes where necessary. For example, see Code 1.4.1.

Code 1.4.1. This is an example of a coding environment, where a certain code of a computer package is provided below.

filename.mpl

```
# some code here
# some other lines of code

# A lot more omitted
# See actual file for full code
```

All relevant code to this thesis is included on the supplementary CD, under the same Chapter headings where they appear in the thesis itself.

In this thesis, the computer packages used are Matlab® , Maple® , Mathematica® and R. Different packages have their own advantages and some are more suitable for a specific task than others. For example, both Maple® and Mathematica® are used for symbolic computation, while Matlab® is better for numerical computation. R is particularly popular for statisticians as it has lots of statistical functionality and it is a free package. We note that Mathematica® also has the ability to interact with R through its `RLink` package from verison 9, but Maple® does not have this feature yet.

The machine specifications used for computations are given below:

```
Laptop:
Model: Alienware M15X       CPU: i7 940X
OS:    Windows 7 SP1 X64    RAM: 4GB

SMSAS Server: (Emmy)
Model: N/A                  CPU: Xeon E5540
OS:    Red Hat 6.5 X64      RAM: 64GB

Computer packages:
Maple 18.01 X64
Mathematica 10.0.1 X64
R 3.1.1 X64
Matlab R2014b X64
```

The two main computer packages used for the thesis are Mathematica® and Maple® . Source codes and relevant worksheets are attached electronically in the supplementary CD. We will discuss briefly below some advantages of each packages.

Mathematica® is a commercial package and is capable of doing both symbolic and numerical calculations. We used Mathematica® in all chapters. Mathematica® 's built–in parallel feature is more powerful than Maple® and it also has four built–in global optimization algorithms. For example, we can utilize all 32 available cores on the server `Emmy` to speed up calculations using syntax `Parallelize`. The global optimization algorithms available are Nelder–Mead (Nelder and Mead, 1965), differential evolution (Storn and Price, 1995),

simulated annealing (Kirkpatrick *et al.*, 1983) and random search (Zhigljavsky, 1991).

Maple® is another commercial package and is also capable of doing both symbolic and numerical calculations. The PLUR matrix decomposition (Corless and Jeffrey, 1997) which we use extensively is conveniently available in Maple® but is not available in Mathematica® . Maple® is also able to reparameterise complicated expressions such as non–rational terms involving exponentials. However, Maple® does not come with any built–in global optimization functions and these have to be purchased as an additional addon package. Parallelisation is also available but one has to rewrite most of the code and the set up of parallel computing is not an easy task. Maple® may give the wrong rank of a symbolic matrix when some entries are non–rational, see Example 4.2.3.

Chapter 2

Mixture Models in Closed Populations

In this chapter, we focus on the estimation of the size of closed populations. We give a definition of what it is meant by "closed" in Section 2.1, and we describe the sampling method of capture–recapture. In Section 2.2 we provide illustrative closed population real data sets. We also discuss the mixture models that have been used and introduce our new mixture model, which is a mixture of the binomial distribution and the logistic normal binomial distribution. Methods for model fitting are given in Section 2.3 and we present the fitted results in Section 2.4. We conclude this chapter in Section 2.5 and present a list of computer files used in Section 2.6.

2.1 What is a Closed Population?

In ecology, a closed system assumes that there is no birth, no death, and no migrations. These assumptions enable us to estimate the total population size of the animals under the study.

Ecologists often take regular samples from the same area at different sam-

pling occasions; typically populations are sampled on a yearly basis. The captured individuals are marked with a unique tag for identification and released back into the population. This is known as capture–recapture.

2.2 Data and Models

Let K be the total number of sampling occasions. We denote by f_j the number of animals that have been captured j times for $j = 0, \ldots, K$. Table 2.1 presents 18 real datasets summarized in Morgan and Ridout (2008) and Morgan and Ridout (2009). We also consider some larger datasets appearing in Dorazio and Royle (2003), for the North American Breeding Bird Survey (BBS) in Table 2.2, where the length of studies is $K = 50$ for all years from 1997 to 2001. More details on the implementation of the BBS survey are described in Robbins *et al.* (1986) and Boulinier *et al.* (1998).

Having seen the form of the data, we now describe how we use statistical models to model the data and therefore estimate the population size of interest. Consider a study of $K = 8$ sampling occasions. Figure 2.1 illustrates how we set up the closed population model. For each record f_j, denotes the number

# of animals captured j times, f_j	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
Cell probabilities p_j	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8

Figure 2.1: An illustration of the setup of the closed population model

of animals captured j times in the whole study period K, correspondingly p_j denotes the probability associated with f_j. Note that in fact we do not observe all individuals in the population. If we let f_0 be the number of individuals that are not seen during the study, then p_0 is the probability that an animal is not seen in the whole period of the study. It follows that $\sum_{j=0}^{K} p_j = 1$ is a natural constraint. We model $\{f_j, j = 0, \ldots, K\}$ as a multinomial distribution. The

Data	f_j = Number of individuals captured j times																		K	Source
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		
Voles1	29	15	15	16	27	–	–	–	–	–	–	–	–	–	–	–	–	–	5	(Pollock *et al.*, 1990, p. 16)
Voles2	18	15	8	6	5	–	–	–	–	–	–	–	–	–	–	–	–	–	5	Nichols *et al.* (1984)
Chipmunks	14	13	18	12	7	5	1	1	0	1	0	0	0	–	–	–	–	–	13	(Mares *et al.*, 1981, p. 321)
Hares	25	22	13	5	1	2	–	–	–	–	–	–	–	–	–	–	–	–	6	(Otis *et al.*, 1978, p. 36)
Skinks	56	19	28	18	24	14	9	–	–	–	–	–	–	–	–	–	–	–	7	Phillpot (2000)
Rabbits	43	16	8	6	0	2	1	0	0	0	0	0	0	0	0	0	0	0	18	(Edwards and Eberhardt, 1967, p. 90)
Squirrels	23	14	9	6	8	7	3	0	2	0	0	–	–	–	–	–	–	–	11	(Nixon *et al.*, 1967, p. 99)
DeermiceA	9	9	10	8	8	7	–	–	–	–	–	–	–	–	–	–	–	–	6	(Otis *et al.*, 1978, p. 87)
DeermiceB	34	20	28	15	13	–	–	–	–	–	–	–	–	–	–	–	–	–	5	(Otis *et al.*, 1978, p. 92)
Housemice	2	64	40	31	16	13	5	1	0	1	–	–	–	–	–	–	–	–	10	(Otis *et al.*, 1978, p. 63)
Pocketmice	16	15	6	5	5	5	3	–	–	–	–	–	–	–	–	–	–	–	7	(Otis *et al.*, 1978, p. 43)
Woodmice	71	59	41	39	20	26	19	12	9	5	8	4	9	2	1	3	3	3	21	(Tanton, 1965, p. 16)
TaxicabsA	142	81	49	7	3	1	0	0	0	0	–	–	–	–	–	–	–	–	10	(Carothers, 1973, p. 127)
TaxicabsB	104	67	51	12	6	1	0	0	0	0	–	–	–	–	–	–	–	–	10	(Carothers, 1973, p. 127)
Golftees	46	28	21	13	23	14	6	6	11	–	–	–	–	–	–	–	–	–	8	Borchers *et al.* (2002)
Link1	84	54	36	21	–	–	–	–	–	–	–	–	–	–	–	–	–	–	4	Morgan and Ridout (2009)
Link2	32	20	14	11	9	8	6	–	–	–	–	–	–	–	–	–	–	–	7	Morgan and Ridout (2009)
Link3	679	531	379	272	198	143	99	67	46	32	22	14	9	5	3	1	0	0	20	Morgan and Ridout (2009)

Table 2.1: Examples of closed population data sets

f_j	B1997	B1998	B1999	B2000	B2001
1	14	14	11	14	15
2	10	9	12	7	7
3	6	7	10	5	5
4	3	4	4	5	2
5	3	2	4	5	5
6	6	4	1	4	6
7	2	4	4	3	3
8	3	0	2	4	2
9	2	4	3	1	1
10	1	2	3	1	2
11	1	0	0	2	4
12	1	1	2	1	4
13	0	0	4	1	0
14	3	3	1	3	1
15	2	0	1	3	0
16	0	3	0	1	2
17	3	1	1	1	3
18	1	0	1	0	1
19	1	0	1	2	1
20	1	1	2	0	1
21	0	1	0	0	1
22	0	0	1	0	0
23	1	1	0	1	0
24	2	1	0	0	0
25	0	2	0	1	1
26	0	0	0	1	1
27	1	1	0	0	0
28	0	0	0	1	1
29	0	1	0	0	0
30	0	1	1	0	1
31	0	0	3	1	0
32	0	0	0	0	0
33	0	0	0	0	0
34	0	0	0	0	0
35	0	0	0	1	0
36	0	0	0	0	1

Table 2.2: North American Breeding Bird Survey (BBS) data for year 1997, 1998, 1999, 2000 and 2001. For all datasets, we have $K = 50$ and $f_j = 0, j \geq 37$.

full likelihood is given by

$$L(\boldsymbol{\theta}, N; \boldsymbol{f_j}) = N! \prod_{j=0}^{K} \frac{p_j(\boldsymbol{\theta})^{f_j}}{f_j!}, \tag{2.1}$$

where $\boldsymbol{\theta}$ denotes the model parameters and $N = f_0 + f_1 + \cdots + f_K$, is the unknown population size.

We model the capture probabilities p_j via different probability distributions. Morgan and Ridout (2008) presented a mixture model of a binomial and a beta–binomial distribution, and compared the performance of this model to the performances of a single binomial distribution, a mixture of two binomial distributions and a beta–binomial distribution. The paper also considers a standard logistic–normal binomial distribution (Coull and Agresti, 1999). Link (2003) also used a log–gamma distribution to model the capture probabilities. In this chapter, we introduce a mixture of a binomial and a logistic–normal binomial model. We will discuss each of these models below.

Mixture of binomial distributions: The model was considered in Morgan and Ridout (2008) and the capture probabilities $\{p_j\}$s are given by

$$p_j = \binom{K}{j} \sum_{c=1}^{C} w_c \phi_c^j (1-\phi_c)^{K-j}, \tag{2.2}$$

for $j = 0, \ldots, K$, where ϕ_c is the constant recapture probability from the cth component, C is the number of mixture components and w_c is the weight of the cth component. We require that $\sum_{c=1}^{C} w_c = 1$. For example, the capture probabilities from a single binomial distribution are

$$p_j = \binom{K}{j} \phi^j (1-\phi)^{K-j}, \tag{2.3}$$

or a mixture of two binomial distributions gives

$$p_j = \binom{K}{j} \left\{ w_1 \phi_1^j (1-\phi_1)^{K-j} + (1-w_1)\phi_2^j (1-\phi_2)^{K-j} \right\}. \tag{2.4}$$

Beta–Binomial distribution: This was proposed by Burnham (1972) for use in this context. The capture probabilities are given by

$$p_j = \binom{K}{j} \frac{B(\alpha + j, \beta - j + K)}{B(\alpha, \beta)}, \tag{2.5}$$

where B denotes the beta function, also known as the beta integral (Whittaker and Watson, 1990),

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\, \mathrm{d}t \tag{2.6}$$

with $\alpha > 0, \beta > 0$. Alternative reparameterisations of the Beta function are given in Read *et al.* (2006),

$$B(\alpha, \beta) = \frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!} \quad \text{if } \alpha, \beta \text{ are integers,} \tag{2.7a}$$

$$= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{2.7b}$$

$$= \sum_{n=0}^{\infty} \binom{n-\beta}{n} \frac{1}{(\alpha+n)}. \tag{2.7c}$$

Note that if we reparameterise using $\mu = \alpha/(\alpha+\beta), \theta = 1/(\alpha+\beta)$, we get the capture probabilities as

$$p_j = \binom{K}{j} \frac{\prod_{r=0}^{j-1}(\mu + r\theta) \prod_{r=0}^{K-j-1}(1-\mu+r\theta)}{\prod_{r=0}^{K-1}(1+r\theta)}, \quad \text{for} \quad 0 < \mu < 1, \theta > 0, \tag{2.8}$$

as given in Morgan and Ridout (2008). In the above, when $\theta \to 0$, we get the binomial distribution back. So this model is a compound distribution of the binomial distribution with the probability parameter following a

beta–binomial distribution.

Binomial Beta–Binomial distribution: This is the new model proposed in Morgan and Ridout (2008). The capture probabilities are given by

$$p_j = \binom{K}{j} \left\{ w\phi^j (1-\phi)^{K-j} + (1-w)\frac{B(\alpha+j, \beta-j+K)}{B(\alpha,\beta)} \right\}, \quad (2.9)$$

where ϕ is the constant recapture probability from the binomial mixture component. This model is a mixture of a binomial distribution with a beta–binomial distribution.

Log–Gamma distribution: The capture probabilities are given by

$$p_j = \binom{K}{j} \int_0^\infty \Big\{ \exp(-x) \Big\}^j \Big\{ 1 - \exp(-x) \Big\}^{K-j} f(x)\, \mathrm{d}x, \quad (2.10)$$

for $j = 0, \ldots, K$, where $f(x)$ is a gamma density function,

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x), \quad \text{for } x \geq 0, \quad (2.11)$$

with shape parameter $\alpha > 0$ and rate parameter $\beta > 0$. This model is a compound distribution of the binomial distribution with the logarithm of the probability parameter following a gamma distribution.

Logistic–Normal Binomial distribution: (Coull and Agresti, 1999) The capture probabilities are given by

$$p_j = \binom{K}{j} \int_{-\infty}^\infty \left\{ \frac{1}{1+\exp(-x)} \right\}^j \left\{ 1 - \frac{1}{1+\exp(-x)} \right\}^{K-j} f(x)\, \mathrm{d}x, \quad (2.12)$$

for $j = 0, \ldots, K$, where $f(x)$ is a normal density function,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2\sigma^2}(x-\mu)^2 \right\}, \quad (2.13)$$

with mean μ and variance $\sigma^2 > 0$.

Binomial Logistic–Normal Binomial distribution: The capture probabilities are given by

$$p_j = \binom{K}{j}\Big[w\phi^j(1-\phi)^{K-j} + (1-w)pj_{\text{LNB}}\Big], \tag{2.14}$$

where pj_{LNB} are the capture probabilities from the logistic normal binomial distribution.

We will focus on the performance of the new model, the binomial logistic–normal binomial model, and compare the fitted result with the existing models.

2.3 Model Fitting

Table 2.3 summarises the numbers of model parameters and gives short abbreviations for all models considered in previous section. In this section, we will discuss the fitting of these mixture models.

Model	Parameters	No. Parameters	Abbreviation
Binomial	f_0, ϕ	2	`Bin`
Two Binomial	f_0, ϕ_1, ϕ_2, w	4	`TwoBin`
Beta Binomial	f_0, α, β	3	`BetaBin`
Binomial Beta Binomial	$f_0, \phi, \alpha, \beta, w$	5	`BBB`
Log Gamma	f_0, α, β	3	`LogGamma`
Logistic Normal Binomial	f_0, μ, σ	3	`LNB`
Binomial Logistic Normal Binomial	$f_0, \phi, \mu, \sigma, w$	5	`BinLNB`

Table 2.3: Summary of closed population models

The fitting of models `Bin` and `TwoBin` is straightforward for most computer programs. Models `BetaBin` and `BBB` may be difficult to fit due to the presence of the beta function. Sometimes the beta function can cause overflow/underflow, which means that the values are too big/small for a computer to cope with. This could be fixed by choosing an appropriate reparameterisation in the formulation of the model. These four models can be fitted using

R package `estimateN.R`, which was developed in Ridout (2007) and was also used in Morgan and Ridout (2009).

Models such as `LogGamma`, `LNB` and `BinLNB` are the most difficult models amongst those considered to fit due to the presence of the integrals. Link (2003) has fitted the models `LogGamma` and `LNB`. A possible solution is to use numerical approximation methods, such as the Gaussian quadrature (Anderson, 1965).

Table 2.4 summarises a list of numerical integration methods using Gaussian quadrature. We will only focus on the Gauss–Hermite quadrature for

Table 2.4: Rules of Gaussian Quadrature (Stroud and Secrest, 1966)

$$\int_a^b w(x)f(x)\,\mathrm{d}x \approx \sum_{i=1}^{n} w_i f(x_i)$$

Quadrature	Interval	Weights	Polynomial
Legendre	$[-1,1]$	1	$P_n(x)$
Chebyshev (1st)	$(-1,1)$	$\frac{1}{\sqrt{1-x^2}}$	$T_n(x)$
Chebyshev (2nd)	$[-1,1]$	$\sqrt{1-x^2}$	$U_n(x)$
Laguerre	$[0,\infty)$	$x^\alpha \mathrm{e}^{-x}$	$L_n^\alpha(x)$
Hermite	$(-\infty,\infty)$	e^{-x^2}	$H_n(x)$
Jacobi	$(-1,1)$	$(1-x)^\alpha(1+x)^\beta$	$J_n^{\alpha,\beta}(x)$
Lobatto	$[-1,1]$	1	$P'_{n-1}(x)$
Kronrod	$[-1,1]$	1	$P_n(x)$

the integrand in the logistic normal binomial model and the Gauss–Laguerre quadrature for the integrand in the log gamma model here. Details on the orthogonal polynomials can be found in Davis and Rabinowitz (1984).

Consider the integrand in Equation (2.12) from the model `LNB`,

$$\begin{aligned} g(x) &= \left\{\frac{1}{1+\exp(-x)}\right\}^{j}\left\{1-\frac{1}{1+\exp(-x)}\right\}^{K-j}\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \\ &\propto \left\{\frac{1}{1+\exp(-x)}\right\}^{j}\left\{1-\frac{1}{1+\exp(-x)}\right\}^{K-j}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}. \end{aligned} \tag{2.15}$$

By comparing the form of $g(x)$ with the "Weights" column and the limits of the integration in Table 2.4, we found that $g(x)$ can be estimated using the Gauss–Hermite quadrature (Liu and Pierce, 1994) through Equation (2.16)

$$\int_{-\infty}^{\infty} h(x)\exp(-x^2)\,\mathrm{d}x \approx \sum_{i=1}^{n} w_i h(x_i), \tag{2.16}$$

with $h(x)$ to be determined, In Equation (2.16), the x_i are the zeros of the nth order Hermite polynomial $H_n(x)$,

$$H_n(x) = n!\sum_{k=0}^{n/2}\frac{(-1)^{n/2-k}}{(2k)!(n/2-k)!}(2x)^{2k}, \tag{2.17}$$

with corresponding weights defined as

$$w_i = \frac{2^{n-1}n!\sqrt{\pi}}{n^2\left[H_{n-1}(x_i)\right]^2}, \tag{2.18}$$

see Greenwood and Miller (1984). Abramowitz and Stegun (1964) gives a table of the first 20 sets of (x_i, w_i), compiled from Salzer *et al.* (1952). A more general Gaussian Hermite method is discussed in Shao *et al.* (1964). Quadrature points for the normal distribution can also be obtained from the R package `npmlreg`.

To find $h(x)$ in Equation (2.16), we substitute $y = (x-\mu)/\sigma$ into $g(x)$ to

give

$$g(y) \propto \left\{\frac{1}{1+\exp(-\mu-y\sigma)}\right\}^{j}\left\{1-\frac{1}{1+\exp(-\mu-y\sigma)}\right\}^{K-j} \mathrm{e}^{-y^2/2}. \quad (2.19)$$

We then substitute $y^2/2$ with x^2 to get,

$$g(x) \propto \left\{\frac{1}{1+\exp(-\mu-x\sqrt{2}\sigma)}\right\}^{j}\left\{1-\frac{1}{1+\exp(-\mu-x\sqrt{2}\sigma)}\right\}^{K-j} \mathrm{e}^{-x^2}, \quad (2.20)$$

which is of the form $h(x)\exp(-x^2)$. Now we can estimate Equation (2.12) by

$$p_j \approx \frac{1}{\sqrt{\pi}}\binom{K}{j}\sum_{i=1}^{n} w_i h(x_i), \quad (2.21)$$

where

$$h(x_i) = \left\{\frac{1}{1+\exp(-\mu-x_i\sqrt{2}\sigma)}\right\}^{j}\left\{1-\frac{1}{1+\exp(-\mu-x_i\sqrt{2}\sigma)}\right\}^{K-j}. \quad (2.22)$$

Alternative methods for numerical integration for the logistic–normal models are discussed in Crouch and Spiegelman (1990) and González *et al.* (2006).

For the log–gamma model, the limits of the integration are different from the limits of the integration in the logistic normal model. We therefore use a different type of Gaussian quadrature, the Gauss Laguerre quadrature (Salzer and Zucker, 1949), to estimate the probabilities in Equation (2.10) using

$$p_j \approx \frac{\binom{K}{j}}{\Gamma(\alpha)}\sum_{i=1}^{n} w_i\left[\left\{\exp\left(\frac{-x_i}{\beta}\right)\right\}^{j}\left\{1-\exp\left(\frac{-x_i}{\beta}\right)\right\}^{K-j} x^{\alpha-1}\right], \quad (2.23)$$

with x_is from the Laguerre polynomial

$$L_n(x) = \sum_{k=0}^{n}\frac{(-1)^k}{k!}\binom{n}{k}x^k, \quad (2.24)$$

and their corresponding weights w_is given by

$$w_i = \frac{x_i}{(n+1)^2\{L_{n+1}(x_i)\}^2}. \tag{2.25}$$

Mathar (2013) tabulates more values of both Gaussian Hermite and Gauss Laguerre weights $\{w_i\}$ and their nodes $\{x_i\}$ up to the 128th order ($n = 128$).

2.4 Results

The maximized log likelihood values ℓ are given in Table 2.5 for data sets given in Table 2.1 and Table 2.2. We note that for the North American Breeding Bird Survey datasets (B1997 – B2001), the best models selected based on maximized log likelihood values are the binomial beta–binomial model (`BBB`) and the binomial logistic normal binomial model (`BinLNB`).

We observe that based on the maximum–likelihood, we would select the models `BBB` and `BinLNB` except for the datasets `link1` and `voles2`, although for the `taxicabsA` and `taxicabsB` data sets, the best models selected are models `TwoBin` and `BinLNB`. The maximum–likelihood values from the two models agree up to 15 decimal places for these two data sets. However, if we look at the maximum likelihood values, the values from the model `TwoBin` and model `BinLNB` are identical up to four decimal places. The maximized likelihood values from the model `BBB` and the model `BinLNB` are very close too. The model `BinLNB` brings in a mixture of a discrete distribution and a continuous distribution.

We can compute the Akaike information criterion (Akaike, 1974) using

$$\text{AIC} = 2 \times \text{Number of parameters} - 2\ell. \tag{2.26}$$

The AIC values give us a model selection criterion, choosing the model with

	Bin	BetaBin	TwoBin	BBB	LogGamma	LNB	BinLNB
B1997	−196.8355	−41.5726	−54.4547	**−40.0532**	−41.7650	−42.4542	−40.1553
B1998	−256.4519	−46.8344	−68.7470	**−44.6311**	−46.9151	−47.3307	−45.0024
B1999	−252.1742	−48.0674	−80.0469	−45.7694	−47.8241	−47.5532	**−45.5621**
B2000	−246.0340	−43.7093	−75.0355	−43.3785	−43.7485	−44.1453	**−43.3282**
B2001	−254.7147	−47.3726	−84.8437	−46.3905	−47.5536	−48.3889	**−46.2575**
chipmunks	−21.0846	−17.1870	−17.0486	**−16.4826**	−17.1454	−17.0755	−16.4945
deermiceA	−21.5592	−11.0248	−10.7529	**−10.6674**	−11.0052	−11.0285	−10.6733
deermiceB	−24.2268	−17.4438	−15.1165	**−15.1165**	−17.4938	−17.4893	**−15.1165**
golftees	−104.9259	−22.5908	−27.1997	−22.0395	−22.4887	−22.6270	**−21.8180**
hares	−15.3309	−12.5538	−10.4344	−10.3461	−12.4957	−12.4177	**−10.3430**
housemice	−44.4259	−43.5509	−39.5377	−39.4741	−43.5147	−43.4475	**−39.4675**
link1	−23.8504	−10.8366	**−10.4795**	−10.4795	−10.7939	−10.7548	−10.4795
link2	−52.6465	−14.9952	−15.8073	**−14.5515**	−14.9199	−14.8096	−14.5645
link3	−878.0716	−47.0122	−105.8911	**−45.7392**	−46.2469	−47.2343	−45.8539
pocketmice	−33.1489	−14.0375	−12.3665	**−12.3345**	−13.9840	−13.7879	−12.3365
rabbits	−21.2877	−14.5515	−14.2984	**−14.2473**	−14.5482	−14.6383	−14.2504
skinks	−86.7133	−22.3548	−23.0394	**−18.5731**	−22.3097	−22.3309	−18.6083
squirrels	−39.3835	−19.2447	−17.8176	−17.7551	−19.3026	−19.4287	**−17.7545**
taxicabsA	−16.9453	−16.4386	**−16.3354**	−16.3354	−16.4406	−16.9453	**−16.3354**
taxicabsB	−17.2623	−16.3017	**−15.8790**	−15.8790	−16.3152	−17.2623	**−15.8790**
voles1	−52.0176	−12.0856	−12.4100	**−11.4702**	−14.3959	−11.8880	−11.5154
voles2	−16.7297	−10.1400	**−9.1918**	−9.1918	−10.1040	−10.0454	−9.1918
woodmice	−357.2700	−47.4526	−87.7337	−45.3368	−46.8905	−45.9236	**−45.0592**

Table 2.5: Maximized log likelihood for closed population datasets. All values are rounded to 4 decimal places. The bold numbers are the largest maximum log likelihood values for each data set.

the lowest AIC. Table 2.6 gives the AIC values for each dataset. The smallest

	Bin	BetaBin	TwoBin	BBB	LogGamma	LNB	BinLNB
B1997	397.6711	**89.1452**	116.9094	90.1064	89.5300	90.9085	90.3105
B1998	516.9038	99.6687	145.4940	**99.2622**	99.8301	100.6614	100.0048
B1999	508.3485	102.1348	168.0938	101.5388	101.6483	**101.1064**	101.1242
B2000	496.0681	**93.4187**	158.0709	96.7570	93.4970	94.2906	96.6564
B2001	513.4295	**100.7453**	177.6874	102.7810	101.1072	102.7778	102.5150
chipmunks	46.1693	40.3741	42.0971	42.9653	40.2908	**40.1511**	42.9890
deermiceA	47.1185	28.0495	29.5058	31.3348	**28.0104**	28.0569	31.3466
deermiceB	52.4537	40.8875	**38.2330**	40.2330	40.9877	40.9785	40.2330
golftees	213.8518	51.1816	62.3993	54.0790	**50.9774**	51.2539	53.6360
hares	34.6618	31.1075	**28.8688**	30.6922	30.9913	30.8354	30.6860
housemice	92.8519	93.1018	**87.0755**	88.9482	93.0294	92.8951	88.9350
link1	51.7008	27.6731	28.9590	30.9590	27.5879	**27.5097**	30.9590
link2	109.2930	35.9903	39.6147	39.1030	35.8399	**35.6191**	39.1291
link3	1760.1432	100.0244	219.7821	101.4785	**98.4937**	100.4686	101.7078
pocketmice	70.2979	34.0749	**32.7330**	34.6691	33.9679	33.5757	34.6730
rabbits	46.5755	35.1031	36.5969	38.4946	**35.0965**	35.2765	38.5008
skinks	177.4267	50.7096	54.0789	**47.1461**	50.6194	50.6618	47.2166
squirrels	82.7671	44.4893	**43.6353**	45.5101	44.6052	44.8573	45.5090
taxicabsA	**37.8906**	38.8772	40.6707	42.6707	38.8813	39.8906	42.6707
taxicabsB	**38.5247**	38.6034	39.7580	41.7580	38.6303	40.5247	41.7580
voles1	108.0352	30.1712	32.8200	32.9404	34.7919	**29.7760**	33.0307
voles2	37.4594	26.2801	26.3837	28.3837	26.2080	**26.0908**	28.3837
woodmice	718.5400	100.9053	183.4674	100.6737	99.7810	**97.8471**	100.1184

Table 2.6: AIC values for closed population data sets. All values are rounded to 4 decimal places. The bold numbers are the smallest for each data sets.

AIC value in each row is given in bold.

For the `link3` dataset, we show the plot of the densities of the capture probabilities in Figure 2.2. As a comparison, we have an extra curve describing the capture probability from the model `BinLNB` compared with Figure 1

from Morgan and Ridout (2009). The authors noted that the heterogeneity of

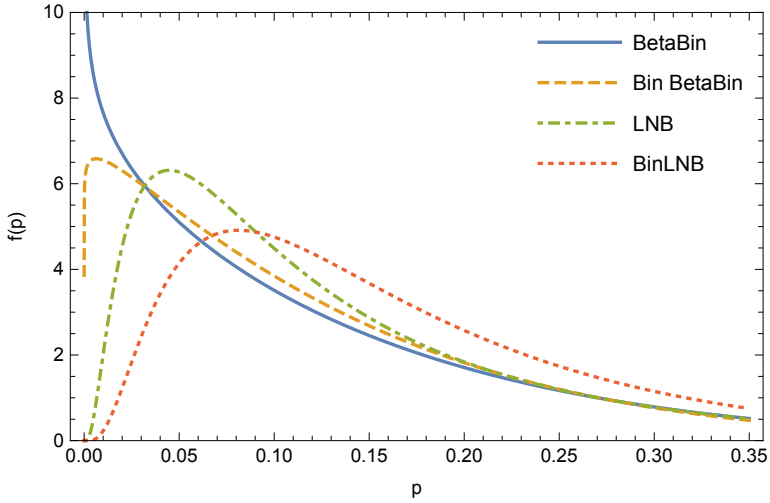

Figure 2.2: A comparison of model fits for the `link3` dataset. Here we show how the four models describe the capture probability near zero.

the capture probability near the origin is handled differently by the different models. The continuous distribution always has density zero at the origin and this is a severe restriction. For the `link3` dataset, the model `BBB` is a compromise between the model `BetaBin` and the model `LNB`. We observe that due to the normal distribution component in the new model, `BinLNB` shifts more away from zero compared with the other models.

2.5 Conclusion

We have seen that for a closed–population, if we assume all individuals have the same recapture probability, we can model the population through a multinomial distribution. By having different probability distributions for the $\{f_j\}$s, we obtain different models.

We presented the data and a range of mixture models. We have shown how we estimate the integrals of the logistic–normal distribution and the log–gamma distribution using different Gauss Quadratures. We have presented our fitted results and compared the performance of the model `BinLNB` with the rest of the models.

To summarise, the new model `BinLNB` performs equally well compared to the model `BBB` and has the potential to be extended to include behavioural effects as well as individuals random effects using the logistic link. In order to fit the new model to individual data, to consider the possible value of time and behaviour effects, it was necessary to develop an individual–based likelihood. Checks of this, not shown here, were satisfied when it was possible also to compare with likelihoods when the time and behavioural effects were not present.

2.6 Code & List of files

Table 2.7 gives a list of supplementary files for this chapter on the CD.

File Name	Description
`chap2.pdf`	Mathematica® file used for this chapter.
`maxlltab.pdf`	Table 2.5 with the maximum likelihood values in bold and red.
`AICtable.pdf`	Table 2.6 with the minimum AIC values in bold and red.
`mletable.pdf`	Maximum likelihood estimates obtained from the optimization.
`chap2_data_ms_2000.xlxs`	All results in Excel sheet, which can be imported to Mathematica® .

Table 2.7: List of files on Chapter 2

Chapter 3

Mixture Models in Open Populations

3.1 Introduction

We have seen in Chapter 2 that for closed populations, we do not allow individuals to enter or leave the population. In contrast, we now consider open populations where individuals are allowed to enter and leave the population as well as migrate. For example, individuals can enter the population through birth and leave the population through death.

In the sections below, we discuss the mixture structures for the following three types of models, the capture–recapture model in Section 3.2, the capture–recovery model in Section 3.3 and the capture–recapture–recovery model in Section 3.4. For each of the three models, we explain the form of the data and the model setup using examples. This chapter sets the background for all the models we consider in later chapters.

3.2 capture–recapture

3.2.1 Background

In this section, we focus on two types of models for describing capture–recapture data, namely the Cormack–Jolly–Seber Model (CJS), see Lebreton *et al.* (1992), and the Jolly–Seber Model (JS), after Jolly (1965) and Seber (1965). The CJS model first appeared in Cormack (1964), and only allows time–varying estimates of survival probabilities and recapture probabilities. The JS model is more general so that it also allows the estimation of population size and the number of new individuals entering the population (White *et al.*, 1982).

We will first explain how capture–recapture data are collected and what the data look like, in Section 3.2.2. We will then focus on the CJS model in Section 3.2.3 and the JS model in Section 3.2.4.

3.2.2 Data

In Chapter 2, our interest is on the population size, we do not model any individual effects, therefore, the summarised data on the frequencies being recaptured are sufficient. In open population, we are more interested in individuals. To allow for groups of individuals, we need to have the individual capture histories.

In a capture–recapture study of length K, the animals are captured, marked and released at every sampling occasion $j = 1, \ldots, K$, typically measured in years. For an animal i that was captured or recaptured at time j, we record its capture history $\mathbf{CH}_i$ with entries $x_{ij} = 1$ for a capture, or $x_{ij} = 0$ otherwise.

These records form a capture history matrix $\mathbf{CH}$. The dimension of this matrix is given by $n \times K$, where n is the number of distinct animals recorded during the study. The ith row of $\mathbf{CH}$ represents the complete capture history

of individual i, denoted by $\mathbf{CH}_i$. We define f_i to be the first capture time and ℓ_i to be the last capture time of animal i.

For example, suppose we have the capture history for an individual i,

$$\mathbf{CH}_i = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}. \tag{3.1}$$

The length of $\mathbf{CH}_i$ is 7, so the length of the study is $K = 7$. The animal was first caught at time $f_i = 2$ and last seen at time $\ell_i = 6$. It was not seen at time $j = 3$, but was recaptured at times $j = 4, 5$ and $j = 6$. It was not seen at the last sampling occasion $j = 7$, when the study ends.

Example 3.2.1 (Capture–recapture data in m–array format)**.** Sometimes we also present the data in a more compact way as an m–array. Consider the European Dipper *(Cinclus cinclus)* dataset from Lebreton *et al.* (1992), the individual capture histories are given below

```
111 100 111 110
110 110 100 110
111 100 111 100
110 100 100 101
100 110 100 110
101 100
```

The length of study is $K = 3$. There are 22 capture histories. We can summaries them as given in Table 3.1. Among the 22 dippers, 11 were recaptured

Year Released	Number Released	First recapture		
		Year 2	Year 3	Never recaptured
1	22	11 (11-)	2 (101)	9 (100)
2	11		4 (111)	7 (110)
3	6			6 (--0)

Table 3.1: Capture–recapture data in an m–array form. European Dipper dataset from Lebreton *et al.* (1992), with capture histories in the brackets.

in the second year with capture history '11–', with '–' indicating either a '0' or a '1'. Among these 11 dippers, 4 were recaptured in the third year with capture history '111'. In the third year of study, year 3, we have released 6 dippers with capture history '– – 1', with '–' indicating either a '0' or a '1'.

The detailed algorithm to obtain the m–array format from the capture–history matrix **CH** can be found in Lebreton *et al.* (1992). It is also available in the `R` package `Rcapture` using the function `descriptive`, which is discussed in Baillargeon and Rivest (2007). □

3.2.3 Cormack–Jolly–Seber Model

We start by considering the basic Cormack–Jolly–Seber (CJS) model, where both the survival probabilities and recapture probabilities vary with time.

Let ϕ_j denote the probability of survival between times j and $j+1$, and let p_j denote the probability of recapture at time j. For a capture history with $K = 7$, we would have the following setup:

$$1 \xrightarrow{\phi_1} \underset{p_2}{2} \xrightarrow{\phi_2} \underset{p_3}{3} \xrightarrow{\phi_3} \underset{p_4}{4} \xrightarrow{\phi_4} \underset{p_5}{5} \xrightarrow{\phi_5} \underset{p_6}{6} \xrightarrow{\phi_6} \underset{p_7}{7}.$$

Numbers 1 to 7 denote the sampling occasions. In this example, we can think of them as being $j = 1, 2, \ldots, 7$. The survival probabilities $\{\phi_j\}_1^6$ are labelled above the arrow as each ϕ_j is defined to be the survival probability between times j and $j+1$. Beneath each j, we have p_j, which is defined as the probability of recapture at time j.

Using the capture history from Equation (3.1), we have

$$0 \to 1 \xrightarrow{\phi_2} \underset{1-p_3}{0} \xrightarrow{\phi_3} \underset{p_4}{1} \xrightarrow{\phi_4} \underset{p_5}{1} \xrightarrow{\phi_5} \underset{p_6}{1} \to 0.$$

We replace j with the capture history at time j and change the labels of ϕ_j and

p_j accordingly. For example, the individual was not captured at time $j = 3$, and so we change the label beneath position 3 from p_3 to $1-p_3$. The individual was last seen at time $j = 6$, since at position 7 we have a record of 0. So we no longer have ϕ_6 between position 6 and 7. It follows that the probability of this capture history $\Pr(\mathbf{CH}_i)$ is given by

$$\phi_2(1-p_3)\phi_3 p_4\phi_4 p_5\phi_5 p_6(1-\phi_6) + \phi_2(1-p_3)\phi_3 p_4\phi_4 p_5\phi_5 p_6\phi_6(1-p_7). \quad (3.2)$$

As the fate of the animal is not known after occasion 6, the first term of the above probability represents the probability that the animal did not survive after time 6, and the second term represents the probability that the animal had survived after time 6, but it was not recaptured at time 7.

The probability of a capture history from the CJS model can be written in the following form (Pledger *et al.*, 2003),

$$\Pr(\mathbf{CH}_i) = \sum_{d=\ell_i}^{K} \left\{ \left(\prod_{j=f_i}^{d-1} \phi_j \right) (1-\phi_d) \left(\prod_{j=f_i+1}^{d} p_j^{x_{ij}} (1-p_j)^{1-x_{ij}} \right) \right\}. \quad (3.3)$$

Note. We define the empty products $\prod_{j=f_i}^{d-1}$ and $\prod_{j=f_i+1}^{d}$ to be 1 when $f_i > d-1$ and $f_i + 1 > d$ respectively. We also set $\phi_K = 0$ since we do not have any information on survival after the study between times K and $K+1$, and $p_1 = 1$ as recapture can only start from time 2. There are $K-1$ survival probabilities and $K-1$ recapture probabilities. The CJS model has a total of $2K-2$ parameters.

Example 3.2.2 (CJS Model)**.** Consider a study with length $K = 3$. For completeness, we consider all $2^3 - 1 = 7$ *observable* capture histories. Using

Equation (3.3), we then obtain the following probabilities,

$$\begin{bmatrix} \Pr(001) \\ \Pr(010) \\ \Pr(011) \\ \Pr(100) \\ \Pr(101) \\ \Pr(110) \\ \Pr(111) \end{bmatrix} = \begin{bmatrix} 1 \\ 1-\phi_2+\phi_2(1-p_3) \\ \phi_2 p_3 \\ 1-\phi_1+(1-\phi_2)\phi_1(1-p_2)+\phi_1(1-p_2)\phi_2(1-p_3) \\ \phi_1(1-p_2)\phi_2 p_3 \\ (1-\phi_2)\phi_1 p_2+\phi_1 p_2\phi_2(1-p_3) \\ \phi_1 p_2\phi_2 p_3 \end{bmatrix}. \quad (3.4)$$

We do not consider the probability Pr(000) in the CJS model as we do not observe it. The CJS model is constructed to be conditional on the first capture.

Among the 7 probabilities above, we observe that

$$\Pr(001) = 1 \quad (3.5a)$$

$$\Pr(010) + \Pr(011) = 1 \quad (3.5b)$$

$$\Pr(100) + \Pr(101) + \Pr(110) + \Pr(111) = 1. \quad (3.5c)$$

It follows that there are only four *independent* capture histories (with associated probabilities). This means that we only have four pieces of usable information from $K = 3$ years of data. In practice, if there are missing capture histories, we could have less usable information. □

Figure 3.1 shows a tree diagram of the CJS capture histories. At each sampling occasion, there are 2^{j-1} terms that correspond to animals which were first caught in the same year $j, j = 1, \ldots, K$ and the probabilities of those capture history terms sum to 1. In general for a K–year study, we have $2^K - 1$ observable capture histories. But only $2^K - K - 1$ independent capture histories can be used to estimate the parameters. This can be interpreted as there

Figure 3.1: Illustration of capture histories for CJS models of capture–recapture data

being 2^K ways of arranging 0s and 1s less K constraints and less 1 invalid way of arrange them (the one correspond to a string of 0s). In other words, among the $2^K - 1$ observable histories, there are K constraints, one for each year of first capture. Hence, there are only $2^K - K - 1$ independent capture histories.

The CJS model assumes that all animals have the same $\{\phi_j\}$ and $\{p_j\}$. We can relax this assumption by introducing different groups of animals. Within each group, the animals share the same survival and recapture probabilities. Pledger (2000) noted that it is not necessary to assume animals have actual different groups. Often the groups are essentially an artefact to detect heterogeneity in the data if it is present.

To describe heterogeneity of both capture and survival, it is assumed that there are C classes with probabilities $\{w_c\}$ of an animal being in class c. Each animal in class c has survival probability $\{\phi_{jc}\}$ and capture probability $\{p_{jc}\}$ for $c = 1, \ldots, C$ and $j = 1, \ldots, K$. The mixture version of Equation (3.3) was derived in Pledger *et al.* (2003) to give the following probability of a capture history

$$\Pr(\mathbf{CH}_i) = \sum_{d=\ell_i}^{K} \sum_{c=1}^{C} \left\{ w_c \left(\prod_{j=f_i}^{d-1} \phi_{jc} \right) (1 - \phi_{dc}) \left(\prod_{j=f_i+1}^{d} p_{jc}^{x_{ij}} (1 - p_{jc})^{1-x_{ij}} \right) \right\}, \tag{3.6}$$

where $\sum_{c=1}^{C} w_c = 1$ is a proper constraint. For example, we choose to set $w_C = 1 - \sum_{c=1}^{C-1} w_c$.

When heterogeneity is present, we model $\{\phi_{jc}\}$ and $\{p_{jc}\}$ through different link functions $g(\phi_{jc})$ (and/or $g(p_{jc})$), using the notations defined in Notation 3.2.1.

Notation 3.2.1. For the survival probabilities $\{\phi_{jc}\}$, we use μ_ϕ as baseline from group 1, $\tau_{\phi j}$ as a time component and $\eta_{\phi c}$ as a heterogeneous component. Since we have μ_ϕ describing the baseline from group 1, we set up constraints

$\tau_{\phi 1} = 0$ and $\eta_{\phi 1} = 0$. So that $\tau_{\phi j}$ are the differences in the time component between the baseline and time component at j; $\eta_{\phi c}$ are the differences in the heterogeneous component between the baseline and group c.

Similarly for the capture probabilities $\{p_{jc}\}$, we use μ_p as baseline from group 1, τ_{pj} as a time component and η_{pc} as a heterogeneous component, with constraints $\tau_{p2} = 0$ (since the recapture probability starts from time $j = 2$) and $\eta_{p1} = 0$. By convention, we set $\phi_{Kc} = 0$ and $p_{1c} = 1$ for all c. That is, the survival probability after the length of study K is 0 and the capture probability at time 1 is 1 for all groups.

For $K = 4, C = 3$, for example, we can model ϕ_{jc} using a linear link function,

$$\phi_{jc} = \mu_\phi + \tau_{\phi j} + \eta_{\phi c} \tag{3.7a}$$

$$[\phi_{jc}] = \begin{bmatrix} \mu_\phi & \mu_\phi + \eta_{\phi 2} & \mu_\phi + \eta_{\phi 3} \\ \mu_\phi + \tau_{\phi 2} & \mu_\phi + \tau_{\phi 2} + \eta_{\phi 2} & \mu_\phi + \tau_{\phi 2} + \eta_{\phi 3} \\ \mu_\phi + \tau_{\phi 3} & \mu_\phi + \tau_{\phi 3} + \eta_{\phi 2} & \mu_\phi + \tau_{\phi 3} + \eta_{\phi 3} \\ 0 & 0 & 0 \end{bmatrix}, \tag{3.7b}$$

or using a logistic link function,

$$\log\left(\frac{\phi_{jc}}{1-\phi_{jc}}\right) = \mu_\phi + \tau_{\phi j} + \eta_{\phi c} \tag{3.8a}$$

$$[\phi_{jc}] = \begin{bmatrix} \frac{1}{1+\mathrm{e}^{-(\mu_\phi)}} & \frac{1}{1+\mathrm{e}^{-(\mu_\phi+\eta_{\phi 2})}} & \frac{1}{1+\mathrm{e}^{-(\mu_\phi+\eta_{\phi 3})}} \\ \frac{1}{1+\mathrm{e}^{-(\mu_\phi+\tau_{\phi 2})}} & \frac{1}{1+\mathrm{e}^{-(\mu_\phi+\tau_{\phi 2}+\eta_{\phi 2})}} & \frac{1}{1+\mathrm{e}^{-(\mu_\phi+\tau_{\phi 2}+\eta_{\phi 3})}} \\ \frac{1}{1+\mathrm{e}^{-(\mu_\phi+\tau_{\phi 3})}} & \frac{1}{1+\mathrm{e}^{-(\mu_\phi+\tau_{\phi 3}+\eta_{\phi 2})}} & \frac{1}{1+\mathrm{e}^{-(\mu_\phi+\tau_{\phi 3}+\eta_{\phi 3})}} \\ 0 & 0 & 0 \end{bmatrix}. \tag{3.8b}$$

In the matrix $[\phi_{jc}]$, each row describes the time effect and each column describes the heterogeneous effect.

It is also possible to include interactions between the time component and the heterogeneous component in the model by setting

$$\phi_{jc} = \mu_\phi + \tau_{\phi j} + \eta_{\phi c} + (\tau\eta)_{\phi jc}, \tag{3.9}$$

where the matrix $(\tau\eta)_\phi$ takes the following form

$$(\tau\eta)_\phi = \begin{pmatrix} \tau\eta_{\phi 11} & \tau\eta_{\phi 12} & \cdots & \tau\eta_{\phi 1C} \\ \tau\eta_{\phi 21} & \tau\eta_{\phi 22} & \cdots & \tau\eta_{\phi 2C} \\ \vdots & \cdots & \ddots & \vdots \\ \tau\eta_{\phi K1} & \tau\eta_{\phi K2} & \cdots & \tau\eta_{\phi KC} \end{pmatrix}. \tag{3.10}$$

Extra constraints are needed on $(\tau\eta)_\phi$, such as requiring each row and column to sum to zero. For simplicity, we set the first column and last row of $(\tau\eta)_\phi$ to

zero as constraints. Note that using different constraints can result in different point estimates (for example each entry in the matrix ϕ_{jc}). Models with interactions are often more complicated than they should be and have too many parameters for successful model fitting (Pledger *et al.*, 2010). We will discuss this feature at more length in Section 6.4. A similar setup can be used for p_{jc}.

Often we model probabilities using the logistic link function, so that the point estimates are between 0 and 1. When different link functions are used, we have different point estimates of the parameters. In particular, when using the method of maximum–likelihood estimation, we will get different results from using different link functions. The effect of using different link functions will be discussed in more detail in Section 6.3.

Note. Depending on the values taken by K and C, models can have different numbers of parameters. Table 3.2 summarises the numbers of model parameters of the 25 models given in Pledger *et al.* (2003). We use '$\cdot$' for a constant parameter, 't' for a time varying parameter, 'h_C' for a heterogeneous parameter with C classes and '$\times$' for a parameter with interactions.

The square brackets indicate that both parameters are heterogeneous but they share the same $\{w_c\}$s; see Pledger *et al.* (2003).

The authors give the total number of estimable parameters in the last column of Table 1 in Pledger *et al.* (2003) as the column 'Total'. In our table, we present the column 'Total' as the total number of parameters in the model. We will discuss the problem of 'estimable parameters' in Chapter 4 when considering *parameter redundancy*.

3.2.4 Jolly–Seber Model

Pledger *et al.* (2010) considers a full likelihood–based version of the JS model

Model	w_c	ϕ	p	Total
$\{\phi(\cdot), p(\cdot)\}$	0	1	1	2
$\{\phi(\cdot), p(t)\}$	0	1	$K-1$	K
$\{\phi(\cdot), p(h_C)\}$	$C-1$	1	C	$2C$
$\{\phi(\cdot), p(t+h_C)\}$	$C-1$	1	$K+C-2$	$K+2C-2$
$\{\phi(\cdot), p(t\times h_C)\}$	$C-1$	1	$(K-1)C$	KC
$\{\phi(t), p(\cdot)\}$	0	$K-1$	1	K
$\{\phi(t), p(t)\}$	0	$K-1$	$K-1$	$2K-2$
$\{\phi(t), p(h_C)\}$	$C-1$	$K-1$	C	$K+2C-2$
$\{\phi(t), p(t+h_C)\}$	$C-1$	$K-1$	$K+C-2$	$2K+2C-4$
$\{\phi(t), p(t\times h_C)\}$	$C-1$	$K-1$	$(K-1)C$	$KC+K-2$
$\{\phi(h_C), p(\cdot)\}$	$C-1$	C	1	$2C$
$\{\phi(h_C), p(t)\}$	$C-1$	C	$K-1$	$K+2C-2$
$\{[\phi(h), p(h)]_C\}$	$C-1$	C	C	$3C-1$
$\{[\phi(h), p(t+h)]_C\}$	$C-1$	C	$K+C-2$	$K+3C-3$
$\{[\phi(h), p(t\times h)]_C\}$	$C-1$	C	$(K-1)C$	$KC+C-1$
$\{\phi(t+h_C), p(\cdot)\}$	$C-1$	$K+C-2$	1	$K+2C-2$
$\{\phi(t+h_C), p(t)\}$	$C-1$	$K+C-2$	$K-1$	$2K+2C-4$
$\{[\phi(t+h), p(h)]_C\}$	$C-1$	$K+C-2$	C	$K+3C-3$
$\{[\phi(t+h), p(t+h)]_C\}$	$C-1$	$K+C-2$	$K+C-2$	$2K+3C-5$
$\{[\phi(t+h), p(t\times h)]_C\}$	$C-1$	$K+C-2$	$(K-1)C$	$KC+K+C-3$
$\{\phi(t\times h_C), p(\cdot)\}$	$C-1$	$(K-1)C$	1	KC
$\{\phi(t\times h_C), p(t)\}$	$C-1$	$(K-1)C$	$K-1$	$KC+K-2$
$\{[\phi(t\times h), p(h)]_C\}$	$C-1$	$(K-1)C$	C	$KC+C-1$
$\{[\phi(t\times h), p(t+h)]_C\}$	$C-1$	$(K-1)C$	$K+C-2$	$KC+K+C-3$
$\{[\phi(t\times h), p(t\times h)]_C\}$	$C-1$	$(K-1)C$	$(K-1)C$	$2KC-C-1$

Table 3.2: Numbers of model parameters with K samples and C classes for CJS mixture models. For model notations see text or Pledger *et al.* (2003).

from Schwarz and Arnason (1996), based on the idea of an entry parameter developed in Crosbie and Manly (1985). In comparison to the CJS model, which only focuses on survival $\{\phi_j\}$ and recapture probabilities $\{p_j\}$, the JS model allows the estimation of population size N as well as the numbers of new individuals entering the population at different times.

Schwarz and Arnason (1996) suggested a super–population of N individuals, in which each animal is available for capture at least once during the whole study of length K. We use $\beta_0, \beta_1, \ldots, \beta_{K-1}$ to denote the proportions of the N animals which enter the population and become available for capture at sampling times $j = 1, 2, \ldots, K$ respectively. We require that $\sum_0^{K-1} \beta_j = 1$.

Unlike the CJS model, which conditions on first capture, the JS model can use a capture history $\mathbf{CH}_0 = \mathbf{0}$ for an animal that was never caught or seen in the whole study period, with its corresponding probability given by

$$\Pr(\mathbf{CH}_0) = \sum_{b=1}^{K}\sum_{d=b}^{K}\sum_{c=1}^{C}\left\{w_c\beta_{b-1}\left(\prod_{j=b}^{d-1}\phi_{jc}\right)(1-\phi_{dc})\left(\prod_{j=b}^{d}(1-p_{jc})\right)\right\}. \tag{3.11}$$

For an animal i that was seen at least once, the probability of such a capture history is given by

$$\Pr(\mathbf{CH}_i) = \sum_{b=1}^{f_i}\sum_{d=\ell_i}^{K}\sum_{c=1}^{C}\left\{w_c\beta_{b-1}\left(\prod_{j=b}^{d-1}\phi_{jc}\right)(1-\phi_{dc})\left(\prod_{j=b}^{d}p_{jc}^{x_{ij}}(1-p_{jc})^{1-x_{ij}}\right)\right\}. \tag{3.12}$$

Both expressions for $\Pr(\mathbf{CH}_0)$ and $\Pr(\mathbf{CH}_i)$ are given in Pledger *et al.* (2010).

Example 3.2.3 (JS model)**.** Consider a study with length $K = 3$. The unique

capture histories give the following probabilities,

$$\begin{bmatrix} \Pr(000) \\ \Pr(001) \\ \Pr(010) \\ \Pr(011) \\ \Pr(100) \\ \Pr(101) \\ \Pr(110) \\ \Pr(111) \end{bmatrix} = \begin{bmatrix} \beta_1\bar{\phi}_2\bar{p}_2 + \beta_1\phi_2\bar{p}_2\bar{p}_3 + (1-\beta_0-\beta_1)\bar{p}_3 \\ \beta_1\phi_2\bar{p}_2 p_3 + (1-\beta_0-\beta_1)p_3 \\ \beta_1\bar{\phi}_2 p_2 + \beta_1\phi_2 p_2\bar{p}_3 \\ \beta_1\phi_2 p_2 p_3 \\ \beta_0\bar{\phi}_1 + \beta_0\phi_1\bar{\phi}_2\bar{p}_2 + \beta_0\phi_1\phi_2\bar{p}_2\bar{p}_3 \\ \beta_0\phi_1\phi_2\bar{p}_2 p_3 \\ \beta_0\phi_1\bar{\phi}_2 p_2 + \beta_0\phi_1\phi_2 p_2\bar{p}_3 \\ \beta_0\phi_1\phi_2 p_2 p_3 \end{bmatrix}, \tag{3.13}$$

where $\bar{\theta} = 1 - \theta$. □

Note. As with the CJS model, depending on different values taken by K and C, the JS models can also have different numbers of parameters. Table 3.3 summarises the numbers of model parameters given in Pledger *et al.* (2010). The only difference in the number of parameters between a CJS model and its corresponding JS model is given by the parameters $\{\beta_j\}$, which are $K-1$ in number, and the parameter for the total population size N. The same notations for models have been used as in Table 3.2.

3.3 Capture–Recovery

3.3.1 Background

The data in capture–recovery studies are different from those in capture–recapture investigations, as animals are never recaptured alive but instead may be recovered dead.

Finite mixture models can be applied in the same way as for capture–recapture models. Pledger and Schwarz (2002) discussed such models in great detail and fitted models to both real and simulated data.

Model	N	β	w_c	ϕ	p	Total
$\{\beta(t), \phi(\cdot), p(\cdot)\}$	1	$K-1$	0	1	1	$K+2$
$\{\beta(t), \phi(\cdot), p(t)\}$	1	$K-1$	0	1	$K-1$	$2K$
$\{\beta(t), \phi(\cdot), p(h_C)\}$	1	$K-1$	$C-1$	1	C	$K+2C$
$\{\beta(t), \phi(\cdot), p(t+h_C)\}$	1	$K-1$	$C-1$	1	$K+C-2$	$2K+2C-2$
$\{\beta(t), \phi(\cdot), p(t \times h_C)\}$	1	$K-1$	$C-1$	1	$(K-1)C$	$KC+K$
$\{\beta(t), \phi(t), p(\cdot)\}$	1	$K-1$	0	$K-1$	1	$2K$
$\{\beta(t), \phi(t), p(t)\}$	1	$K-1$	0	$K-1$	$K-1$	$3K-2$
$\{\beta(t), \phi(t), p(h_C)\}$	1	$K-1$	$C-1$	$K-1$	C	$2K+2C-2$
$\{\beta(t), \phi(t), p(t+h_C)\}$	1	$K-1$	$C-1$	$K-1$	$K+C-2$	$3K+2C-4$
$\{\beta(t), \phi(t), p(t \times h_C)\}$	1	$K-1$	$C-1$	$K-1$	$(K-1)C$	$KC+2K-2$
$\{\beta(t), \phi(h_C), p(\cdot)\}$	1	$K-1$	$C-1$	C	1	$K+2C$
$\{\beta(t), \phi(h_C), p(t)\}$	1	$K-1$	$C-1$	C	$K-1$	$2K+2C-2$
$\{\beta(t), \phi(h_C), p(h_C)\}$	1	$K-1$	$C-1$	C	C	$K+3C-1$
$\{\beta(t), \phi(h_C), p(t+h_C)\}$	1	$K-1$	$C-1$	C	$K+C-2$	$2K+3C-2$
$\{\beta(t), \phi(h_C), p(t \times h_C)\}$	1	$K-1$	$C-1$	C	$(K-1)C$	$KC+K+C-1$
$\{\beta(t), \phi(t+h_C), p(\cdot)\}$	1	$K-1$	$C-1$	$K+C-2$	1	$2K+2C-2$
$\{\beta(t), \phi(t+h_C), p(t)\}$	1	$K-1$	$C-1$	$K+C-2$	$K-1$	$3K+2C-4$
$\{\beta(t), \phi(t+h_C), p(h_C)\}$	1	$K-1$	$C-1$	$K+C-2$	C	$2K+3C-3$
$\{\beta(t), \phi(t+h_C), p(t+h_C)\}$	1	$K-1$	$C-1$	$K+C-2$	$K+C-2$	$3K+3C-5$
$\{\beta(t), \phi(t+h_C), p(t \times h_C)\}$	1	$K-1$	$C-1$	$K+C-2$	$(K-1)C$	$KC+2K+C-3$
$\{\beta(t), \phi(t \times h_C), p(\cdot)\}$	1	$K-1$	$C-1$	$(K-1)C$	1	$KC+K$
$\{\beta(t), \phi(t \times h_C), p(t)\}$	1	$K-1$	$C-1$	$(K-1)C$	$K-1$	$KC+2K-2$
$\{\beta(t), \phi(t \times h_C), p(h_C)\}$	1	$K-1$	$C-1$	$(K-1)C$	C	$KC+K+C-1$
$\{\beta(t), \phi(t \times h_C), p(t+h_C)\}$	1	$K-1$	$C-1$	$(K-1)C$	$K+C-2$	$KC+2K+C-3$
$\{\beta(t), \phi(t \times h_C), p(t \times h_C)\}$	1	$K-1$	$C-1$	$(K-1)C$	$(K-1)C$	$2KC+K-C-1$

Table 3.3: Numbers of model parameters with K samples and C classes for JS mixture models

3.3.2 Data

Data are collected over K regular intervals. A new animal i is marked at time j and in the capture history matrix $\mathbf{CH}$ we represent this as $\mathbf{CH}_{ij} = 1$, then the animal is released back into the population. We then only record death at time $k > j$ for the same animal as $\mathbf{CH}_{ik} = 2$.

Example 3.3.1. Consider a $K = 4$–year study of capture–recovery: we list all possible records in a capture–recovery matrix as

$$\mathbf{CH} = \begin{bmatrix} 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & 2 \\ 0 & 1 & 2 & 0 \\ 1 & 0 & 0 & 2 \\ 1 & 0 & 2 & 0 \\ 1 & 2 & 0 & 0 \end{bmatrix}. \tag{3.14}$$

For example, $\mathrm{CH}_{32} = 1$ indicates that the animal was marked at sampling occasion $j = 2$ and $\mathrm{CH}_{33} = 2$ tells us the same animal was found and recorded dead at the next sampling occasion $j = 3$. □

For a study of length K, assuming no animals are recovered beyond the end of time K, we have $K(K-1)/2$ unique capture–recovery histories. This is because for individuals captured in year K, there are $K - 1$ years for them to be recaptured either alive or dead. Since in capture–recovery we are only recording dead individuals, there are $K(K-1)/2$ histories.

3.3.3 Model

Let ϕ_j denote the survival probability between time j and time $j + 1$ and let λ_j denote the recovery or reporting probability at time j; it follows that the

capture–recovery probabilities are given by

$$\Pr(\mathbf{CH}_i) = \left(\prod_{j=f_i+1}^{\ell_i - 1} \phi_{j-1} \right) (1 - \phi_{\ell_i - 1}) \lambda_{\ell_i - 1}. \tag{3.15}$$

Example 3.3.2. Applying Equation (3.15) to the data of (3.14), we obtain

$$\begin{bmatrix} \Pr(0012) \\ \Pr(0102) \\ \Pr(0120) \\ \Pr(1002) \\ \Pr(1020) \\ \Pr(1200) \end{bmatrix} = \begin{bmatrix} (1-\phi_3)\lambda_3 \\ \phi_2(1-\phi_3)\lambda_3 \\ (1-\phi_2)\lambda_2 \\ \phi_1\phi_2(1-\phi_3)\lambda_3 \\ \phi_1(1-\phi_2)\lambda_2 \\ (1-\phi_1)\lambda_1 \end{bmatrix}. \tag{3.16}$$

□

A mixture version of Equation (3.15) is given by

$$\Pr(\mathbf{CH}_i) = \sum_{c=1}^{C} w_c \left\{ \left(\prod_{j=f_i+1}^{\ell_i - 1} \phi_{j-1,c} \right) (1 - \phi_{\ell_i - 1, c}) \lambda_{\ell_i - 1, c} \right\}, \tag{3.17}$$

where we assume there are C different groups and the probabilities $\{w_c\}$ of being in group c. We also set $w_C = 1 - \sum_{c=1}^{C-1} w_c$.

3.4 Capture–Recapture–Recovery

3.4.1 Background

In a capture–recapture–recovery study, we are interested in the survival probabilities $\{\phi_j\}$, the recapture probabilities $\{p_j\}$ and the recovery probabilities $\{\lambda_j\}$, where all parameters are time dependent for example.

3.4.2 Data

Different from capture–recapture and capture–recovery, capture–recapture–recovery studies keep records of individuals that are resighted both alive or dead. Animals are marked and released back into the population after initial capture. Then for a live recapture at time of study j for an individual i, we record $\mathbf{CH}_{ij} = 1$, or for a dead recovery, we record $\mathbf{CH}_{ij} = 2$.

Consider a full capture–recapture–recovery history from a study with length $K = 3$, providing the following capture–recapture–recovery history matrix,

$$\mathbf{CH} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 2 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 0 \end{bmatrix}. \tag{3.18}$$

Consider the history $\mathbf{CH}_{10j} = [1, 1, 2]^T$, which we have not encountered before in capture–recapture or in capture–recovery. We can set up the diagram below for this particular history,

$$1 \xrightarrow{\phi_1} \underset{p_2}{1} \xrightarrow{1-\phi_2} \underset{\lambda_2}{2}.$$

Recall that we define the survival probability ϕ_j between time j and $j + 1$.

Therefore we label ϕ_1 between the history entries 1 and 1 as it survived and $1-\phi_2$ between the history entries 1 and 2 as it was recovered dead. We use p_2 to denote it was recaptured at time 2 and λ_2 to denote it was dead between time 2 and time 3 (it did not survive to time 3). We then multiply the terms to get the probability for this history as $\phi_1 p_2(1-\phi_2)\lambda_2$.

We will now consider a longer history from $K = 7$ for example. The diagram for the capture history

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 2 \end{bmatrix}$$

can be expressed by

$$1 \xrightarrow{\phi_1} \underset{1-p_2}{0} \xrightarrow{\phi_2} \underset{1-p_3}{0} \xrightarrow{\phi_3} \underset{p_4}{1} \xrightarrow{\phi_4} \underset{1-p_5}{0} \xrightarrow{\phi_5} \underset{1-p_6}{0} \xrightarrow{1-\phi_6} \underset{\lambda_6}{2},$$

it follows that its probability is

$$\phi_1(1-p_2)\phi_2(1-p_3)\phi_3 p_4 \phi_4(1-p_5)\phi_5(1-p_6)(1-\phi_6)\lambda_6.$$

For any capture–recapture–recovery history, the diagram can be set up by writing ϕ_j if it survived between time j and time $j+1$ and $1-\phi_j$ between time j and time $j+1$ if it was recovered at time $j+1$. We write p_j under each $\mathbf{CH}_{ij} = 1$ and $1-p_j$ under each $\mathbf{CH}_{ij} = 0$. We write λ_j under each recovery record where $\mathbf{CH}_{ij} = 2$. Then we multiply each of these terms to get the corresponding probability. Using such a system, we can determine the probabilities graphically. A general expression for the model probabilities is given in the next section.

3.4.3 Model

The probability $\Pr(\mathbf{CH}_i)$ from a capture–recapture–recovery history is then given by

$$\begin{cases} \sum_{d=\ell_i}^{K} \left\{ \left(\prod_{j=f_i+1}^{d} \phi_{j-1} p_j^{x_{ij}} (1-p_j)^{1-x_{ij}} \right) (1-\phi_d)(1-\lambda_d) \right\}, & x_{i\ell_i} = 1; \\ \left\{ \prod_{j=f_i+1}^{\ell_i - 1} \phi_{j-1} p_j^{x_{ij}} (1-p_j)^{1-x_{ij}} \right\} (1-\phi_{\ell_i - 1})\lambda_{\ell_i - 1}, & x_{i\ell_i} = 2. \end{cases} \tag{3.19}$$

We use ℓ_i to denote the last time the individual i was seen (either alive or dead). If an animal i was last seen alive, then we have $x_{i\ell_i} = 1$, or if it was last seen dead, then we have $x_{i\ell_i} = 2$.

Example 3.4.1. For a non–mixture version, a study with $K = 3$ will give

$$\begin{bmatrix} \Pr(001) \\ \Pr(010) \\ \Pr(011) \\ \Pr(012) \\ \Pr(100) \\ \Pr(101) \\ \Pr(102) \\ \Pr(110) \\ \Pr(111) \\ \Pr(112) \\ \Pr(120) \end{bmatrix} = \begin{bmatrix} 1 \\ \bar{\phi}_2 \bar{\lambda}_2 + \phi_2 \bar{p}_3 \\ \phi_2 p_3 \\ \bar{\phi}_2 \lambda_2 \\ \bar{\phi}_1 \bar{\lambda}_1 + \phi_1 \bar{p}_2 \bar{\phi}_2 \bar{\lambda}_2 + \phi_1 \bar{p}_2 \phi_2 \bar{p}_3 \\ \phi_1 \bar{p}_2 \phi_2 p_3 \\ \phi_1 \bar{p}_2 \bar{\phi}_2 \lambda_2 \\ \phi_1 p_2 \bar{\phi}_2 \bar{\lambda}_2 + \phi_1 p_2 \phi_2 \bar{p}_3 \\ \phi_1 p_2 \phi_2 p_3 \\ \phi_1 p_2 \bar{\phi}_2 \lambda_2 \\ \bar{\phi}_1 \lambda_1 \end{bmatrix}, \tag{3.20}$$

where $\bar{\theta} = 1 - \theta$. □

For a mixture version of this, let $\{\phi_{jc}\}, \{p_{jc}\}$ and $\{\lambda_{jc}\}$ be the survival probabilities, recapture probabilities and reporting probabilities respectively

from group c, $\Pr(\mathbf{CH}_i)$ is then given by

$$\begin{cases} \sum_{c=1}^{C} \sum_{d=\ell_i}^{K} \left[\left\{ w_c \prod_{j=f_i+1}^{d} \phi_{j-1,c} p_{jc}^{x_{ij}} (1-p_{jc})^{1-x_{ij}} \right\} (1-\phi_{dc})(1-\lambda_{dc}) \right], & x_{i\ell_i} = 1; \\ \sum_{c=1}^{C} w_c \left[\left\{ \prod_{j=f_i+1}^{\ell_i-1} \phi_{j-1,c} p_{jc}^{x_{ij}} (1-p_{jc})^{1-x_{ij}} \right\} (1-\phi_{\ell_i-1,c}) \lambda_{\ell_i-1,c} \right], & x_{i\ell_i} = 2. \end{cases} \tag{3.21}$$

We require that $w_C = 1 - \sum_{c=1}^{C-1} w_c$.

Parameter counting is not given here since we can make any one of the recapture probability, survival probability and/or the recovery probability heterogeneous, time–varying and/or with interactions between the two effects. We will only give a couple of these examples later in Chapter 5 as an illustration of the complexity of parameter redundancy problems.

3.5 Code & List of files

In this section, we list all relevant files on the supplementary CD in Table 3.4. Details on using the Maple® package `MixtureModelforCaptureRecapture.mla` are given in Appendix A and details on using the Mathematica® package `MixtureModelforCaptureRecapture.m` are given in Appendix B.

File Name	Description
`Chapter_3.r`	R script file for Example 3.2.1
`Chapter_3.mw`	Maple® worksheet
`Chapter_3.nb`	Mathematica® notebook
`Chapter_3_R.pdf`	PDF copy of R script
`Chapter_3_Maple.pdf`	PDF copy of Maple® worksheet
`Chapter_3_Mathematica.pdf`	PDF copy of Mathematica® notebook

Table 3.4: List of files on Chapter 3

Code 3.5.1. Here we give the R code that was used in Example 3.2.1 for obtaining the m–array.

`Chapter_3.r`

```
# Example 3.2.1
# R code for Chapter 3
```

```
# Last Updated: 08 March 2015

# Need to install the package 'Rcapture':
install.packages("Rcapture",dep=TRUE)

# load the library
library(Rcapture)

# input dipper data
dipper <- matrix(
    c(
        1,1,1,  1,0,0,  1,1,1,  1,1,0,
        1,1,0,  1,1,0,  1,0,0,  1,1,0,
        1,1,1,  1,0,0,  1,1,1,  1,0,0,
        1,1,0,  1,0,0,  1,0,0,  1,0,1,
        1,0,0,  1,1,0,  1,0,0,  1,1,0,
        1,0,1,  1,0,0
    ),
    nrow = 22,
    ncol = 3,
    byrow = TRUE
)

# obtain a summary of dipper data
summ <- descriptive(dipper)
summ

# obtain the m-array
summ$m

```

For more details on this package, see Baillargeon and Rivest (2007).

Code 3.5.2. This section gives the Maple® code used in this Chapter.

Chapter_3.mw

```
# Example 3.2.2
    myCHcc(3,M);                      # returns capture histories
    mykappa(3,1,CC,tvarphi,tvarp);    # returns probabilities
    mykappa(3,1,tvarphi,tvarp);       # 'CC' (CJS) is by default,
                                      # so can be omitted
# Example 3.2.3
    myCHcc(3,M);                      # returns capture histories
    mykappa(3,1,JS,tvarphi,tvarp);    # returns probabilities

# Example 3.3.2
    myCHcr(4,M);                      # returns capture histories
    mykappa(4,1,CR,tvarphi,tvarlambda); # returns probabilities

# Example 3.4.1
    myCHcrr(3,M);
    mykappa(3,1,CRR,tvarphi,tvarp,tvarlambda);
```

Code 3.5.3. This section gives the Mathematica® code used in this Chapter.

Chapter_3.nb

```
Example 3.2.2
    (* Get Capture Histories *)
    MatrixForm@(CH = myCH["CJS"][3])
```

```
    (* Obtain Probabilities *)
    MatrixForm@(
        mykappa[3, 1, CH, tvarphi -> True,
        tvarp -> True, logit -> False]
    )

Example 3.2.3
    (* Get Capture Histories *)
    MatrixForm@(CH = myCH["JS"][3])
    (* Obtain Probabilities *)
    MatrixForm@(
        mykappa[3, 1, CH, tvarphi -> True,
        tvarp -> True, logit -> False]
    )

Example 3.3.2
    (* Get Capture Histories *)
    MatrixForm@(CH = myCH["CR"][4])
    (* Get Capture Histories *)
    MatrixForm@(
        mykappa[4, 1, CH, tvarphi -> True,
        tvarlambda -> True, logit -> False]
    )

Example 3.4.1
    (* Get Capture Histories *)
    MatrixForm@(CH = myCH["CRR"][3])
    (* Obtain Probabilities *)
    MatrixForm@(
        mykappa[3, 1, CH, tvarphi -> True, tvarp -> True,
        tvarlambda -> True, logit -> False]
    )
```

Chapter 4

Parameter–Redundancy in Open Mixture Models For capture–recapture

In Chapter 3, we have seen that mixture models can have many model parameters. Of interest is whether we can estimate all model parameters given a specific number of years of data, and the choice of number of mixture components to have in the model. Pledger (2000) explored closed population mixture models and a simple rule has been developed in Holzmann *et al.* (2006) to determine whether or not it is possible to distinguish between different models for closed populations. Whether we can apply the same rule to the parameter–redundancy of open population mixture models is explored in this chapter.

We present a motivating example in Section 4.1 to introduce the concept of parameter–redundancy. Then we discuss existing methods of detecting parameter–redundancy and their limitations in Section 4.2. We introduce the term *exhaustive summary* which uniquely determines model structure, and show how it can be used to determine the parameter–redundancy results. We review the symbolic method (Cole *et al.*, 2010) and the hybrid

symbolic–numerical (SN) method (Choquet and Cole, 2012). We compare how the symbolic method can be executed in different computer packages such as Maple® and Mathematica® .

We present our findings of parameter–redundancy results for small values of K and C using the SN method in Section 4.3. Here K is length of the open population study and C is the number of mixture components we want to include in the model. To explore relationships further in order to develop the rules for mixture models in open populations, we need a *simple exhaustive summary*, which is simpler than the original exhaustive summary but carries the same information. The simpler exhaustive summary can either have fewer terms or a simpler structure than the original exhaustive summary. We give the theorems that are needed to develop and prove results for simpler exhaustive summaries for the next chapter in Section 4.4. We give a conclusion of the work of this chapter in Section 4.5 and list the computer codes used for this chapter in Section 4.6.

4.1 What is parameter–redundancy?

In simple words, this is a question of whether we can estimate all of the parameters in a particular model using the method of maximum likelihood estimation. To get a better idea of what parameter–redundancy is, we will start by looking at a simple example below.

Example 4.1.1 (CJS model continued)**.** Recall Example 3.2.2, the right hand side of Equation (3.4) can be simplified to obtain the following expression for

the capture history probabilities,

$$\begin{bmatrix} 1 \\ 1-\boldsymbol{\phi_2 p_3} \\ \boldsymbol{\phi_2 p_3} \\ 1-\phi_1 p_2-\phi_1\boldsymbol{\phi_2 p_3}+\phi_1\boldsymbol{\phi_2} p_2\boldsymbol{p_3} \\ \phi_1\boldsymbol{\phi_2 p_3}(1-p_2) \\ \phi_1 p_2-\phi_1\boldsymbol{\phi_2} p_2\boldsymbol{p_3} \\ \phi_1\boldsymbol{\phi_2} p_2\boldsymbol{p_3} \end{bmatrix}. \tag{4.1}$$

By inspection, we cannot separate parameters ϕ_2 and p_3. These two parameters always appear in the model as a single product. This tells us that we cannot estimate either one of them individually, but in fact we can estimate their product.

One way to get around this problem is to reparameterise the product as a single parameter, for example letting $x_0 = \phi_2 p_3$. In practice, it is more common to set one of them equal to 1. For instance, we set $p_3 = 1$, so that the interpretation of the estimate of ϕ_2 is no longer the survival probability at time 2, but the product of ϕ_2 and p_3. Both methods allow the estimation of the product, but effectively we now have one fewer parameter in the model. □

For any model $\boldsymbol{M}$ with parameters $\boldsymbol{\theta}$ in the parameter space Ω, then the model is parameter redundant if $\boldsymbol{M}(\boldsymbol{\theta})$ can be written in terms of the elements of the function $f(\boldsymbol{\theta}) \in \Omega_\beta$, in which Ω_β has dimension $\dim(\boldsymbol{\beta}) < \dim(\boldsymbol{\theta})$, see Catchpole and Morgan (1997). In Example 4.1.1, the model $\boldsymbol{M}(\boldsymbol{\theta})$ with $\boldsymbol{\theta} = [\phi_1, \phi_2, p_2, p_3]^T$ is parameter redundant, while the model with $\boldsymbol{\beta} = [\phi_1, p_2, \phi_2 p_3]^T$ is not parameter redundant.

Holzmann *et al.* (2006) have shown that for closed population mixture

models, see Pledger (2000) for example, it is necessary to have $K \geq 2C$ to ensure that we can distinguish between different models in a specified family, where K is the number of samples and C is the number of mixture components in the model. In the web supplementary material of Pledger *et al.* (2010), the authors noted that

> Finding exact rules for these more complex open–population models would be a difficult and time–consuming exercise, and the rules are unlikely to be simple.

This is precisely the focus of this chapter and the next chapter, from the prospective of parameter–redundancy. We shall come back to this later.

4.2 Methods for Detecting parameter–redundancy

4.2.1 Background

The main approach we discuss here was developed in Catchpole and Morgan (1997) and Catchpole *et al.* (1998). The method requires the calculation of a derivative matrix symbolically, and its symbolic rank. A comprehensive symbolic approach to determining the parametric structure of models was presented in Cole *et al.* (2010), followed by a hybrid symbolic–numerical method detailed in Choquet and Cole (2012).

4.2.2 Notation

The quantity that uniquely determines the parametric structure of a model is called an *exhaustive summary*, usually denoted by $\boldsymbol{\kappa}$. In all of our open population models, the exhaustive summary is formed from the set of probabilities of each capture history. For example, the expression in (4.1) is an exhaustive summary for the CJS model when $K = 3$. Methods for detecting parameter–

redundancy based on exhaustive summaries have been developed in Cole *et al.* (2010).

Notation 4.2.1. We define the following notation.

$\boldsymbol{\theta}$ is the original model parameter vector.

$\boldsymbol{s}(\boldsymbol{\theta})$ is a reparameterisation in terms of the original model parameters. Typically, the reparameterisation has fewer terms than the set of original model parameters, so that it is simpler. For instance, see Example 4.1.1. The original model parameters are $\boldsymbol{\theta} = [\phi_1, \phi_2, p_2, p_3]^T$ and the reparameterisation is $\boldsymbol{s}(\boldsymbol{\theta}) = [\phi_1, p_2, \phi_2 p_3]^T$.

n_θ is the number of original parameters: $n_\theta = \text{Dimension}(\boldsymbol{\theta})$.

n_s is the number of parameters in the reparameterisation: $n_s = \text{Dimension}(\boldsymbol{s})$.

$\boldsymbol{\kappa}$ is an exhaustive summary:

- $\boldsymbol{\kappa}(\boldsymbol{\theta})$ is the original exhaustive summary in terms of $\boldsymbol{\theta}$.
- $\boldsymbol{\kappa}(\boldsymbol{s})$ is the exhaustive summary in terms of the reparameterisation $\boldsymbol{s}$.

$\boldsymbol{D}$ is a derivative matrix.

- $\boldsymbol{D}(\boldsymbol{\theta})$ is the derivative matrix obtained by differentiating $\boldsymbol{\kappa}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, i.e. $\boldsymbol{D}(\boldsymbol{\theta}) = \partial\boldsymbol{\kappa}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. If the dimension of $\boldsymbol{\kappa}(\boldsymbol{\theta})$ is $n_{\boldsymbol{\kappa}}$, then the dimension of $\boldsymbol{D}(\boldsymbol{\theta})$ is given by $n_\theta \times n_{\boldsymbol{\kappa}}$.
- $\boldsymbol{D}(\boldsymbol{s})$ is the derivative matrix obtained by differentiating $\boldsymbol{\kappa}(\boldsymbol{s})$ with respect to $\boldsymbol{s}$, i.e. $\boldsymbol{D}(\boldsymbol{s}) = \partial\boldsymbol{\kappa}(\boldsymbol{s})/\partial\boldsymbol{s}$.

r is the rank of a matrix.

- r_θ is the rank of $\boldsymbol{D}(\boldsymbol{\theta})$.
- r_s is the rank of $\boldsymbol{D}(\boldsymbol{s})$.

4.2.3 Symbolic Method

For the symbolic method, we introduce the following theorem from Cole *et al.* (2010).

Theorem 4.2.1 (Testing parameter–redundancy)**.** *If $r_{\boldsymbol{\theta}} = n_{\boldsymbol{\theta}}$, the model is full rank and not parameter–redundant. We can estimate all model parameters. If $r_{\boldsymbol{\theta}} < n_{\boldsymbol{\theta}}$, the model is not full rank and parameter–redundant. There are $r_{\boldsymbol{\theta}}$ estimable parameters. We define the* deficiency *of the model as $d_{\boldsymbol{\theta}} = n_{\boldsymbol{\theta}} - r_{\boldsymbol{\theta}}$.*

For parameter–redundant models, the estimable parameters can be obtained by solving $\boldsymbol{\alpha}(\boldsymbol{\theta})^T \boldsymbol{D}(f(\boldsymbol{\theta})) = \mathbf{0}$, with $\boldsymbol{D}(f(\boldsymbol{\theta}))$ being the matrix of derivatives from differentiating the arbitrary function $f(\boldsymbol{\theta})$ with respect to each of the terms in $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}(\boldsymbol{\theta})$ determining the nullspace of $\boldsymbol{D}(\boldsymbol{\theta})$. This forms a linear system of first–order partial differential equations (PDEs), which has the number of solutions equal to model rank $r_{\boldsymbol{\theta}}$.

For proof see Theorem 2 in Cole *et al.* (2010).

Typically, the steps for the symbolic method for determining parameter–redundancy are carried out as follows:

Step 1. List all possible histories $\mathbf{CH}_i$ from the model.

Step 2. Find the probabilities of each history. These probabilities $\Pr(\mathbf{CH}_i)$ form the original exhaustive summary $\boldsymbol{\kappa}(\boldsymbol{\theta})$.

Step 3. Obtain the derivative matrix $\boldsymbol{D}(\boldsymbol{\theta})$ and calculate the rank r_θ.

We then determine the parameter–redundancy result of the model, based on the values of r_θ and $n_{\boldsymbol{\theta}}$ using Theorem 4.2.1. We will illustrate this using the example below.

Example 4.2.1 (CJS model continued)**.** This example illustrates the symbolic method of testing parameter–redundancy on the CJS model. Consider the

original exhaustive summary in Example 4.1.1, we have

$$\boldsymbol{\kappa}(\boldsymbol{\theta}) = \begin{bmatrix} 1 \\ 1-\phi_2 p_3 \\ \phi_2 p_3 \\ 1-\phi_1 p_2 - \phi_1\phi_2 p_3 + \phi_1\phi_2 p_2 p_3 \\ \phi_1\phi_2 p_3(1-p_2) \\ \phi_1 p_2 - \phi_1\phi_2 p_2 p_3 \\ \phi_1\phi_2 p_2 p_3 \end{bmatrix}. \tag{4.2}$$

We differentiate each of the terms with respect to the elements in the parameter vector $\boldsymbol{\theta} = [p_2, p_3, \phi_1, \phi_2]^T$ in turn to obtain

$$\boldsymbol{D}(\boldsymbol{\theta})^T = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -\phi_2 & 0 & -p_3 \\ 0 & \phi_2 & 0 & p_3 \\ -\phi_1\bar{\phi}_2 - \phi_1\phi_2\bar{p}_3 & -\phi_1\bar{p}_2\phi_2 & -1+\bar{p}_2\bar{\phi}_2 + \bar{p}_2\phi_2\bar{p}_3 & -\phi_1\bar{p}_2 + \phi_1\bar{p}_2\bar{p}_3 \\ -p_3\phi_1\phi_2 & \phi_1\bar{p}_2\phi_2 & \bar{p}_2\phi_2 p_3 & \phi_1\bar{p}_2 p_3 \\ \phi_1\bar{\phi}_2 + \phi_1\phi_2\bar{p}_3 & -\phi_1 p_2\phi_2 & p_2\bar{\phi}_2 + p_2\phi_2\bar{p}_3 & -\phi_1 p_2 + \phi_1 p_2\bar{p}_3 \\ p_3\phi_1\phi_2 & \phi_1 p_2\phi_2 & p_2\phi_2 p_3 & p_2 p_3\phi_1 \end{bmatrix}, \tag{4.3}$$

with rank $r_{\boldsymbol{\theta}} = 3$ using symbolic computer packages like Maple® or Mathematica® . Here again, we have used $\bar{\theta} = 1 - \theta$ for shorthand. We have $r_{\boldsymbol{\theta}} = 3 < n_{\boldsymbol{\theta}} = 4$, so that the model is parameter–redundant with deficiency $d_{\boldsymbol{\theta}} = 4 - 3 = 1$. To find the estimable combinations, we need the nullspace of $\boldsymbol{D}(\boldsymbol{\theta})$. We obtain the single solution

$$\boldsymbol{\alpha}(\boldsymbol{\theta})^T = \begin{bmatrix} 0 & -\dfrac{p_3}{\phi_2} & 0 & 1 \end{bmatrix}. \tag{4.4}$$

Together with

$$\boldsymbol{D}(f(\boldsymbol{\theta})) = \begin{bmatrix} \dfrac{\partial f(p_2, p_3, \phi_1, \phi_2)}{\partial p_2} \\ \dfrac{\partial f(p_2, p_3, \phi_1, \phi_2)}{\partial p_3} \\ \dfrac{\partial f(p_2, p_3, \phi_1, \phi_2)}{\partial \phi_1} \\ \dfrac{\partial f(p_2, p_3, \phi_1, \phi_2)}{\partial \phi_2} \end{bmatrix}, \tag{4.5}$$

we obtain a single partial differential equation to solve,

$$-\frac{p_3}{\phi_2}\frac{\partial f(p_2,p_3,\phi_1,\phi_2)}{\partial p_3}+\frac{\partial f(p_2,p_3,\phi_1,\phi_2)}{\partial \phi_2}=0. \tag{4.6}$$

We find that a possible solution to this system is $F(\phi_1, p_2, \phi_2 p_3)$ from using Maple® , where F is also an arbitrary function. Therefore we conclude that we can estimate ϕ_1, p_2 separately and the product $\phi_2 p_3$. The result agrees with what we observed in Example 4.1.1. □

We note that the PDE in Equation (4.6) is simple and can be solved using the method of auxiliary equations, see Myint-U and Debnath (2007). When mixture components are introduced in the model, the system of PDEs are more complicated. We shall see this in the example below.

Example 4.2.2 (Estimable parameters from a mixture model)**.** We will now show an example from a mixture model where the deficiency is more than 1. Consider the model $\{\phi(t+h_C), p(\cdot)\}$ for $K=3, C=3$. Using the linear link function, the survival probabilities are modelled by

$$\phi_{jc}=\mu_\phi+\tau_{\phi j}+\eta_{\phi c}, \tag{4.7}$$

where μ_ϕ is the baseline from group 1, $\{\tau_{\phi j}\}$ are the time components and $\{\eta_{\phi c}\}$ are the heterogeneous components. Model parameters are given by

$$\boldsymbol{\theta}=[p,\eta_{\phi 2},\eta_{\phi 3},\mu_\phi,\tau_{\phi 2},w_1,w_2]^T.$$

The derivative matrix $\boldsymbol{D}(\boldsymbol{\theta})$ has rank 4. So the model is parameter–redundant with deficiency 3.

We obtain the null space of $\boldsymbol{D}(\boldsymbol{\theta})$ using Maple® , and it is given by

$$\alpha(\boldsymbol{\theta}) = \begin{bmatrix} 0 & 0 & 0 \\ \dfrac{\eta_{\phi 3}-\eta_{\phi 2}}{2w_2} & \dfrac{\eta_{\phi 3}^2}{2w_2(\eta_{\phi 3}-\eta_{\phi 2})} & \dfrac{\eta_{\phi 2}w_2-\eta_{\phi 3}w_1-\eta_{\phi 3}w_2}{w_2(\eta_{\phi 3}-\eta_{\phi 2})} \\ \dfrac{\eta_{\phi 3}-\eta_{\phi 2}}{2w_3} & \dfrac{\eta_{\phi 3}(\eta_{\phi 3}-2\eta_{\phi 2})}{2w_3(\eta_{\phi 3}-\eta_{\phi 2})} & \dfrac{\eta_{\phi 2}w_3+\eta_{\phi 3}w_3-\eta_{\phi 2}}{-w_3(\eta_{\phi 3}-\eta_{\phi 2})} \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \tag{4.8}$$

where $w_3 = 1 - w_1 - w_2$. We now consider the derivative of an arbitrary function $f(\boldsymbol{\theta})$,

$$\boldsymbol{D}(f(\boldsymbol{\theta}))^T = \begin{bmatrix} \dfrac{\partial f(\boldsymbol{\theta})}{\partial p} & \dfrac{\partial f(\boldsymbol{\theta})}{\partial \eta_{\phi 2}} & \dfrac{\partial f(\boldsymbol{\theta})}{\partial \eta_{\phi 3}} & \dfrac{\partial f(\boldsymbol{\theta})}{\partial \mu_\phi} & \dfrac{\partial f(\boldsymbol{\theta})}{\partial \tau_{\phi 2}} & \dfrac{\partial f(\boldsymbol{\theta})}{\partial w_1} & \dfrac{\partial f(\boldsymbol{\theta})}{\partial w_2} \end{bmatrix}. \tag{4.9}$$

Therefore to find the estimable parameters, we need to solve the following system of PDEs

$$\begin{bmatrix} \dfrac{\eta_{\phi 3}-\eta_{\phi 2}}{2w_2}\dfrac{\partial \boldsymbol{f}}{\partial \eta_{\phi 2}} & + & \dfrac{\eta_{\phi 3}-\eta_{\phi 2}}{2w_3}\dfrac{\partial \boldsymbol{f}}{\partial \eta_{\phi 3}} & + & \dfrac{\partial \boldsymbol{f}}{\partial w_2} \\ \dfrac{\eta_{\phi 3}^2}{2w_2(\eta_{\phi 3}-\eta_{\phi 2})}\dfrac{\partial \boldsymbol{f}}{\partial \eta_{\phi 2}} & + & \dfrac{\eta_{\phi 3}(\eta_{\phi 3}-2\eta_{\phi 2})}{2w_3(\eta_{\phi 3}-\eta_{\phi 2})}\dfrac{\partial \boldsymbol{f}}{\partial \eta_{\phi 3}} & + & \dfrac{\partial \boldsymbol{f}}{\partial w_1} \\ \dfrac{\eta_{\phi 2}w_2-\eta_{\phi 3}w_1-\eta_{\phi 3}w_2}{w_2(\eta_{\phi 3}-\eta_{\phi 2})}\dfrac{\partial \boldsymbol{f}}{\partial \eta_{\phi 2}} & + & \dfrac{\eta_{\phi 2}w_3+\eta_{\phi 3}w_3-\eta_{\phi 2}}{-w_3(\eta_{\phi 3}-\eta_{\phi 2})}\dfrac{\partial \boldsymbol{f}}{\partial \eta_{\phi 3}} & + & \dfrac{\partial \boldsymbol{f}}{\partial \mu_\phi} \end{bmatrix} = \boldsymbol{0}. \tag{4.10}$$

Here we have used $\boldsymbol{f} = f(\boldsymbol{\theta})$ for shorthand. Solving the system with Maple® , we obtain the solution as the following:

$$F(p, \tau_{\phi 2}, -w_3\eta_{\phi 3} - w_2\eta_{\phi 2} - \mu_\phi, w_3\eta_{\phi 3}^2 + 2\mu_\phi w_3\eta_{\phi 3} + 2\eta_{\phi 2}\mu_\phi w_2 + \eta_{\phi 2}^2 w_2 + \mu_\phi^2), \tag{4.11}$$

after some algebraic rearrangement. Here we have again used $w_3 = 1 - w_1 - w_2$ for shorthand. This implies that we can only estimate parameters p and $\tau_{\phi 2}$

separately but not the rest of the parameters.

We can further rearrange the above to obtain the estimable parameters as a set given by

$$\left\{p, \tau_{\phi 2}, -\sum_{c=1}^{3} w_c(\mu_\phi + \eta_{\phi c}), \sum_{c=1}^{3} w_c(\mu_\phi + \eta_{\phi c})^2\right\}, \tag{4.12}$$

where $w_3 = 1 - w_1 - w_2$ and $\eta_{\phi 1} = 0$. Being able to rewrite in the form of (4.12) is essential in determining the general parameter–redundancy result. See Section 5.1.1 for more detailed use. □

In theory, the parametric structure of a model can always be determined using the symbolic method. But in practice, this method relies on the ability to determine the rank of a symbolic matrix and where necessary the ability to solve systems of PDEs.

Table 4.1 gives the dimensions of the matrix $\boldsymbol{D}(\boldsymbol{\theta})$ for five different mixture models.

Model	Dimension of $\boldsymbol{D}(\boldsymbol{\theta})$
$\{\phi(t), p(h_C)\}$	$(2^K - 1) \times (K + 2C - 2)$
$\{\phi(t), p(t + h_C)\}$	$(2^K - 1) \times (2K + 2C - 4)$
$\{\phi(t), p(t \times h_C)\}$	$(2^K - 1) \times (KC + K - 2)$
$\{[\phi(t + h), p(h)]_C\}$	$(2^K - 1) \times (K + 3C - 3)$
$\{[\phi(t + h), p(t + h)]_C\}$	$(2^K - 1) \times (2K + 3C - 5)$

Table 4.1: The dimensions of the derivative matrices

The dimension of the derivative matrix is determined by the number of terms in the exhaustive summary (length of $\boldsymbol{\kappa}(\boldsymbol{\theta})$) and the number of parameters n_θ. As we increase K and/or C, we increase the dimension of the symbolic derivative matrix. Both the dimension of the matrix and the structure of the matrix can make it difficult or impossible to calculate the symbolic rank, even with the aid of a computer package, such as Maple® or Mathematica® .

In the worst scenario, an incorrect rank may be produced (Gimenez *et al.*, 2003, p. 715). This was noticed later in Choquet and Cole (2012). Similar cases were found in Cole and Morgan (2010b). A PLUR decomposition (also known as the Turing decomposition) is needed to double check the result. Corless and Jeffrey (1997) have shown that for any rectangular matrix $\boldsymbol{D}$, we can write $\boldsymbol{D} = \boldsymbol{PLUR}$, where $\boldsymbol{P}$ is a permutation matrix, $\boldsymbol{L}$ is a lower triangular matrix, $\boldsymbol{U}$ is an upper triangular matrix and $\boldsymbol{R}$ is a matrix in reduced row echelon form. This decomposition is conveniently available through Maple® using `LUDecomposition(D,output=['P','L','U1','R'])`. It is not available in Mathematica® . As long as the matrices $\boldsymbol{L}, \boldsymbol{U}$ and $\boldsymbol{R}$ are defined, if the determinant of $\boldsymbol{U}$ is always 0, then the matrix $\boldsymbol{D}$ cannot be full rank (Cole *et al.*, 2010). This case is discussed below in Example 4.2.3.

Example 4.2.3 (Failure of Symbolic Computer Packages)**.** Figure 4.1 shows the derivative matrix $\boldsymbol{A}(\boldsymbol{\theta})$ with model parameters

$$\boldsymbol{\theta} = [\phi_1, \phi_2, \phi_3, \phi_4, a_2, a_3, a_4, a_5, m]^T,$$

as given in Gimenez *et al.* (2003). Here we use the same notation as given in Gimenez *et al.* (2003) for the derivative matrix $\boldsymbol{A}(\boldsymbol{\theta})$ instead of our notation $\boldsymbol{D}(\boldsymbol{\theta})$.

The authors determined the rank of $\boldsymbol{A}(\boldsymbol{\theta})$ is 9 using Maple® syntax `Rank(A)` in Maple® 6 and concluded that the model is full rank since $n_{\boldsymbol{\theta}} = 9$.

Choquet and Cole (2012) noted that earlier versions of Maple® can give incorrect ranks due to the fact that they cannot simplify $1/(1+\mathrm{e}^{x})+1/(1+\mathrm{e}^{-x})$ to 1. They further noted that more recent versions of Maple® have this simplification build in and are able to return the correct rank of 8, in the above example.

However, we note that the rank of this matrix from Maple® 18.01 (the most

$$\begin{bmatrix}
\frac{1}{\phi_1} & \frac{1}{\phi_1} & \frac{1}{\phi_1} & \frac{1}{\phi_1} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{\phi_2} & \frac{1}{\phi_2} & \frac{1}{\phi_2} & \frac{1}{\phi_2} & \frac{1}{\phi_2} & \frac{1}{\phi_2} & 0 & 0 & 0 \\
0 & 0 & \frac{1}{\phi_3} & \frac{1}{\phi_3} & 0 & \frac{1}{\phi_3} & \frac{1}{\phi_3} & \frac{1}{\phi_3} & \frac{1}{\phi_3} & 0 \\
0 & 0 & 0 & \frac{1}{\phi_4} & 0 & 0 & \frac{1}{\phi_4} & 0 & \frac{1}{\phi_4} & \frac{1}{\phi_4} \\
\frac{1}{1+e^{a_2+m}} & \frac{-1}{1+e^{-(a_2+m)}} & \frac{-1}{1+e^{-(a_2+m)}} & \frac{-1}{1+e^{-(a_2+m)}} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{1+e^{a_3}} & \frac{-1}{1+e^{-a_3}} & \frac{-1}{1+e^{-a_3}} & \frac{1}{1+e^{a_3+m}} & \frac{-1}{1+e^{-(a_3+m)}} & \frac{-1}{1+e^{-(a_3+m)}} & 0 & 0 & 0 \\
0 & 0 & \frac{1}{1+e^{a_4}} & \frac{-1}{1+e^{-a_4}} & 0 & \frac{1}{1+e^{a_4}} & \frac{-1}{1+e^{-a_4}} & \frac{1}{1+e^{a_4+m}} & \frac{-1}{1+e^{-(a_4+m)}} & 0 \\
0 & 0 & 0 & \frac{1}{1+e^{a_5}} & 0 & 0 & \frac{1}{1+e^{a_5}} & 0 & \frac{1}{1+e^{a_5}} & \frac{1}{1+e^{a_5+m}} \\
\frac{1}{1+e^{a_2+m}} & \frac{-1}{1+e^{-(a_2+m)}} & \frac{-1}{1+e^{-(a_2+m)}} & \frac{-1}{1+e^{-(a_2+m)}} & \frac{1}{1+e^{a_3+m}} & \frac{-1}{1+e^{-(a_3+m)}} & \frac{-1}{1+e^{-(a_2+m)}} & \frac{1}{1+e^{a_4+m}} & \frac{-1}{1+e^{-(a_4+m)}} & \frac{1}{1+e^{a_5+m}}
\end{bmatrix}$$

Figure 4.1: The derivative matrix from Gimenez *et al.* (2003), for which Maple® would give an incorrect answer

recent version at the time of writing) is still 9 if we evaluate the rank directly using `Rank(A)`, see Figure 4.2. We note that if we use `Rank(expand~(A))`, we then get the correct rank of 8, see Figure 4.3. The Maple® function `expand()` expands the expressions of the form e^{a+b} to $e^a e^b$ and expands the expressions of the form e^{a-b} to e^a/e^b. This "simplification" is in general needed in order to get a correct rank of the models involving exponential functions. Another possible workaround is to reparameterise e^x as a single parameter, for example using $e^x = y$. An explanation of such behaviour is given on the `MaplePrime` discussion forum. We also note that Mathematica® 10.0.2 (the most recent version at the time of writing) can give the correct rank using its function `MatrixRank` without any simplification or tricks, see Figure 4.4.

So when using Maple® to work out the rank of a derivative matrix `DD1`, where there are non–rational entries, we should double check the result using both `Rank(DD1)` and `Rank(expand~(DD1))`. To note that `D` in Maple® is a built–in function as a *Differential operator*. So in all our Maple® worksheets, we use `DD1` to represent a derivative matrix.

To show that the matrix $A(\boldsymbol{\theta})$ cannot be full rank, we do a PLUR decomposition of this matrix, and find that the determinant of the upper triangular matrix U is always 0. This contradicts the definition of a PLUR decomposition (Corless and Jeffrey, 1997, Theorem 1). □

The symbolic method can also fail to calculate a rank when the model is too complex or runs out of memory when the dimension of the matrix is too large; see Hunter *et al.* (2009) and Jiang *et al.* (2007). In the next section, we consider a method that can be used to overcome these problems to some extent.

```
> restart:
> interface(version);
                                                
```

```
> # we expand the matrix A
> expand~(A);
> Rank(%);
```

$$\begin{bmatrix}
\frac{1}{\phi_1} & \frac{1}{\phi_1} & \frac{1}{\phi_1} & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{\phi_2} & \frac{1}{\phi_2} & \frac{1}{\phi_2} & \frac{1}{\phi_2} & 0 & 0 & 0 \\
0 & 0 & \frac{1}{\phi_3} & \frac{1}{\phi_3} & \frac{1}{\phi_3} & \frac{1}{\phi_3} & 0 & 0 \\
0 & 0 & 0 & \frac{1}{\phi_4} & 0 & 0 & 0 & \frac{1}{\phi_4} \\
\frac{1}{1+e^{a2}e^m} & -\frac{1}{1+\frac{1}{e^{a2}e^m}} & -\frac{1}{1+\frac{1}{e^{a2}e^m}} & -\frac{1}{1+\frac{1}{e^{a2}e^m}} & \frac{1}{1+e^{a3}e^m} & -\frac{1}{1+\frac{1}{e^{a3}e^m}} & 0 & 0 \\
0 & \frac{1}{1+e^{a3}} & -\frac{1}{1+\frac{1}{e^{a3}}} & -\frac{1}{1+\frac{1}{e^{a3}}} & 0 & \frac{1}{1+e^{a4}} & -\frac{1}{1+\frac{1}{e^{a4}}} & 0 \\
0 & 0 & \frac{1}{1+e^{a4}} & -\frac{1}{1+\frac{1}{e^{a4}}} & 0 & 0 & \frac{1}{1+e^{a5}} & -\frac{1}{1+\frac{1}{e^{a5}}} \\
\frac{1}{1+e^{a2}e^m} & -\frac{1}{1+\frac{1}{e^{a2}e^m}} & -\frac{1}{1+\frac{1}{e^{a2}e^m}} & -\frac{1}{1+\frac{1}{e^{a2}e^m}} & \frac{1}{1+e^{a3}e^m} & -\frac{1}{1+\frac{1}{e^{a3}e^m}} & -\frac{1}{1+\frac{1}{e^{a4}e^m}} & \frac{1}{1+e^{a5}e^m}
\end{bmatrix}$$

This is the rank returned by Maple, not page number -> 8

Figure 4.3: An example of Maple® giving the correct rank, after some necessary simplification.

```
In[1]:= $Version
Out[1]= 10.0 for Microsoft Windows (64-bit) (December 4, 2014)

In[2]:= (A = {
    {1/ϕ1, 1/ϕ1, 1/ϕ1, 1/ϕ1, 0, 0, 0, 0, 0, 0},
    {0, 1/ϕ2, 1/ϕ2, 1/ϕ2, 1/ϕ2, 1/ϕ2, 1/ϕ2, 0, 0, 0},
    {0, 0, 1/ϕ3, 1/ϕ3, 0, 1/ϕ3, 1/ϕ3, 1/ϕ3, 1/ϕ3, 0},
    {0, 0, 0, 1/ϕ4, 0, 0, 1/ϕ4, 0, 1/ϕ4, 1/ϕ4},
    {1/(1 + Exp[a2 + m]), -1/(1 + Exp[-(a2 + m)]),
     -1/(1 + Exp[-(a2 + m)]), -1/(1 + Exp[-(a2 + m)]), 0, 0, 0, 0, 0, 0},
    {0, 1/(1 + Exp[a3]), -1/(1 + Exp[-a3]), -1/(1 + Exp[-a3]), 1/(1 + Exp[a3 + m]),
     -1/(1 + Exp[-(a3 + m)]), -1/(1 + Exp[-(a3 + m)]), 0, 0, 0},
    {0, 0, 1/(1 + Exp[a4]), -1/(1 + Exp[-a4]), 0, 1/(1 + Exp[a4]),
     -1/(1 + Exp[-a4]), 1/(1 + Exp[a4 + m]), -1/(1 + Exp[-(a4 + m)]), 0},
    {0, 0, 0, 1/(1 + Exp[a5]), 0, 0, 1/(1 + Exp[a5]), 0,
     1/(1 + Exp[a5]), 1/(1 + Exp[a5 + m])},
    {1/(1 + Exp[a2 + m]), -1/(1 + Exp[-(a2 + m)]), -1/(1 + Exp[-(a2 + m)]),
     -1/(1 + Exp[-(a2 + m)]), 1/(1 + Exp[a3 + m]),
     -1/(1 + Exp[-(a3 + m)]), -1/(1 + Exp[-(a3 + m)]),
     1/(1 + Exp[a4 + m]), -1/(1 + Exp[-(a4 + m)]), 1/(1 + Exp[a5 + m])}
   }) // MatrixForm
```

Out[2]//MatrixForm=

$$\begin{pmatrix}
\frac{1}{\phi_1} & \frac{1}{\phi_1} & \frac{1}{\phi_1} & \frac{1}{\phi_1} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{\phi_2} & \frac{1}{\phi_2} & \frac{1}{\phi_2} & \frac{1}{\phi_2} & \frac{1}{\phi_2} & \frac{1}{\phi_2} & 0 & 0 & 0 \\
0 & 0 & \frac{1}{\phi_3} & \frac{1}{\phi_3} & 0 & \frac{1}{\phi_3} & \frac{1}{\phi_3} & \frac{1}{\phi_3} & \frac{1}{\phi_3} & 0 \\
0 & 0 & 0 & \frac{1}{\phi_4} & 0 & 0 & \frac{1}{\phi_4} & 0 & \frac{1}{\phi_4} & \frac{1}{\phi_4} \\
\frac{1}{1+e^{a2+m}} & -\frac{1}{1+e^{-a2-m}} & -\frac{1}{1+e^{-a2-m}} & -\frac{1}{1+e^{-a2-m}} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{1+e^{a3}} & -\frac{1}{1+e^{-a3}} & -\frac{1}{1+e^{-a3}} & \frac{1}{1+e^{a3+m}} & -\frac{1}{1+e^{-a3-m}} & -\frac{1}{1+e^{-a3-m}} & 0 & 0 & 0 \\
0 & 0 & \frac{1}{1+e^{a4}} & -\frac{1}{1+e^{-a4}} & 0 & \frac{1}{1+e^{a4}} & -\frac{1}{1+e^{-a4}} & \frac{1}{1+e^{a4+m}} & -\frac{1}{1+e^{-a4-m}} & 0 \\
0 & 0 & 0 & \frac{1}{1+e^{a5}} & 0 & 0 & \frac{1}{1+e^{a5}} & 0 & \frac{1}{1+e^{a5}} & \frac{1}{1+e^{a5+m}} \\
\frac{1}{1+e^{a2+m}} & -\frac{1}{1+e^{-a2-m}} & -\frac{1}{1+e^{-a2-m}} & -\frac{1}{1+e^{-a2-m}} & \frac{1}{1+e^{a3+m}} & -\frac{1}{1+e^{-a3-m}} & -\frac{1}{1+e^{-a3-m}} & \frac{1}{1+e^{a4+m}} & -\frac{1}{1+e^{-a4-m}} & \frac{1}{1+e^{a5+m}}
\end{pmatrix}$$

```
In[3]:= MatrixRank[A]
Out[3]= 8
```

Figure 4.4: An example of Mathematica® giving the correct rank without any simplification.

4.2.4 Hybrid Symbolic–Numerical (SN) Method

An alternative to the symbolic method is to use a numerical approach. Rothenberg (1971) has shown that the Fisher information matrix can be used. Since the Fisher information matrix is the negative of the second derivative of the log–likelihood, we can alternatively make use of the Hessian matrix. Viallefont *et al.* (1998) considers using the Hessian matrix and its rank to determine parameter–redundancy. This numerical approach requires fitting of the model before we can determine whether we can estimate all model parameters. Gimenez *et al.* (2004) has given an example of the using the Hessian matrix when a real dataset is being fitted to the CJS model. The mixture models are particularly difficult to fit as we shall see Chapter 6. We will not discuss this in much detail here.

Instead, Choquet and Cole (2012) consider a hybrid symbolic–numerical method, where the derivative matrix is calculated symbolically, but the rank is evaluated numerically. The hybrid symbolic–numerical method is denoted as the SN method. We implement the SN method using the steps below:

Step 1. List all possible histories $\mathbf{CH}_i$ from the model.

Step 2. Find the probabilities of each history. These probabilities form the original exhaustive summary $\boldsymbol{\kappa}(\boldsymbol{\theta})$.

Step 3. Work out the derivative matrix $\boldsymbol{D}(\boldsymbol{\theta})$.

Step 4. Simulate parameter values for each parameter in $\boldsymbol{\theta}$ and work out the rank numerically.

Choquet and Cole (2012) had shown that five random points in the parameter space are sufficient to determine the model rank using the SN method. The model rank is then taken as the maximum of the five ranks obtained.

For parameter–redundant models, the SN method can also identify individual parameters that can in theory be estimated. We shall see this through Example 4.2.4.

Example 4.2.4 (Finding estimable parameters using the SN method)**.** This example shows how to use hybrid symbolic–numerical method to find which parameters can be estimated separately from the others. Recall that in Example 4.2.1, we obtained the derivative matrix symbolically as

$$\boldsymbol{D}(\boldsymbol{\theta}) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -\phi_2 & 0 & -p_3 \\ 0 & \phi_2 & 0 & p_3 \\ -\phi_1\bar{\phi}_2 - \phi_1\phi_2\bar{p}_3 & -\phi_1\bar{p}_2\phi_2 & -1 + \bar{p}_2\bar{\phi}_2 + \bar{p}_2\phi_2\bar{p}_3 & -\phi_1\bar{p}_2 + \phi_1\bar{p}_2\bar{p}_3 \\ -p_3\phi_1\phi_2 & \phi_1\bar{p}_2\phi_2 & \bar{p}_2\phi_2 p_3 & \phi_1\bar{p}_2 p_3 \\ \phi_1\bar{\phi}_2 + \phi_1\phi_2\bar{p}_3 & -\phi_1 p_2\phi_2 & p_2\bar{\phi}_2 + p_2\phi_2\bar{p}_3 & -\phi_1 p_2 + \phi_1 p_2\bar{p}_3 \\ p_3\phi_1\phi_2 & \phi_1 p_2\phi_2 & p_2\phi_2 p_3 & p_2 p_3\phi_1 \end{bmatrix}. \quad (4.13)$$

We simulate a set of random values for $\boldsymbol{\theta}$ (using the Mathematica® function `RandomReal[1, 4]`), for example

$$\tilde{\boldsymbol{\theta}} = \begin{cases} p_2 &= 0.767829 \\ p_3 &= 0.855950 \\ \phi_1 &= 0.875648 \\ \phi_2 &= 0.587508 \end{cases},$$

and substitute this set of parameter values into $\boldsymbol{D}(\boldsymbol{\theta})$ to obtain

$$\boldsymbol{D}(\tilde{\boldsymbol{\theta}}) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -0.587508 & 0 & -0.85595 \\ 0 & 0.587508 & 0 & 0.85595 \\ -0.435304 & -0.11944 & -0.884583 & -0.174015 \\ -0.440344 & 0.11944 & 0.116754 & 0.174015 \\ 0.435304 & -0.39501 & 0.381705 & -0.575497 \\ 0.440344 & 0.39501 & 0.386124 & 0.575497 \end{bmatrix}.$$

The rank of this derivative matrix is 3. We find that the nullspace of $\boldsymbol{D}(\tilde{\boldsymbol{\theta}})$ is given by the single solution

$$\alpha(\tilde{\boldsymbol{\theta}}) = \begin{bmatrix} 6.82887 \times 10^{-17} \\ -0.824472 \\ -8.53956 \times 10^{-17} \\ 0.565902 \end{bmatrix} \approx \begin{bmatrix} 0 \\ -0.824472 \\ 0 \\ 0.565902 \end{bmatrix}.$$

Recall that the derivative matrix $\boldsymbol{D}(\boldsymbol{\theta})$ was obtained by differentiating $\boldsymbol{\kappa}(\boldsymbol{\theta})$ with respect to the elements of the parameter vector $\boldsymbol{\theta} = [p_2, p_3, \phi_1, \phi_2]^T$ in order. The positions of the zeros (the first and the third) in $\alpha(\tilde{\boldsymbol{\theta}})$ indicate that parameters p_2 and ϕ_1 can be estimated but parameters p_3 and ϕ_2 cannot be estimated separately. □

This hybrid approach is useful when the partial differential equation systems are difficult or impossible to solve. However, for parameter–redundant models, it can only identify which parameters can be estimated separately, and it cannot give the combinations of parameters which can be estimated. To obtain the full set of estimable parameters in such a case, we still require solving the PDEs.

4.3 Results Using the SN Method

Using the SN method, we can determine the model ranks for any values of K and C for mixture models much faster than using the symbolic method. For different dimensions of K and C, we can produce tables of model ranks and therefore model deficiencies as shown in Table 4.2 and Table 4.3 for those small values of K and C shown. Ranks from models with larger values of K and C can be obtained in the same way. See the files listed in Section 4.6 for all tables.

Recall that at the beginning of this chapter we have seen that Holzmann

et al. (2006) developed a rule of $K \geq 2C$, to ensure closed population mixture models are identifiable. We will present this discriminatory rule as a dashed line at $K = 2C$ where appropriate in all tables for comparison.

Table 4.2 shows the parameter–redundancy result for the capture–recapture model $\{[\phi(h), p(h)]_C\}$ for $K = 3, 4, \ldots, 10$ and $C = 2, 3, \ldots, 10$. In this model, we consider heterogeneity in both the survival and the recapture probabilities. In the table, the solid line divides the table into two parts. To the left of the solid line, the model is parameter redundant and to the right of the solid line, the model is full rank.

In Table 4.2(a), the model ranks are given. The ranks are the numbers of parameters that can be estimated. For the parameter–redundant region (to the left of the solid line), the sequence 3,6,10,15,21,28 follows the expression $K(K-1)/2$ and for the full rank region (to the right of the solid line), the sequence 5,8,11,...,29 follows the expression $3C - 1$, which is the number of model parameters. An informed guess for this model to be full rank is therefore $K(K-1)/2 \geq 3C - 1$.

For this model in particular, using the rule $K \geq 2C$ given in Holzmann *et al.* (2006) for closed population mixture models, we will consider all models to the left of the dashed to be parameter redundant (unidentifiable). But in fact, the models between the solid line and the dashed line in the table are also full rank. For example, at $K = 8, C = 6$, the using rule of $K \geq 2C$ determines the model to be parameter redundant but using the rule of $K(K-1)/2 \geq 3C-1$, the model will be full rank.

Table 4.3 shows the parameter–redundancy result for the model $\{\phi(t), p(t \times h_C)\}$ for $K = 3, 4, \ldots, 10$ and $C = 2, 3, \ldots, 10$. In this model, both the survival and the recapture probabilities are time varying. The capture probabilities p_{jc} also have interaction terms between time and heterogeneous groups.

(a) Model ranks based on the linear link. The same result is obtained using the logistic link function.

	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$	$K=8$	$K=9$	$K=10$
$C=2$	3	5	5	5	5	5	5	5
$C=3$	3	6	8	8	8	8	8	8
$C=4$	3	6	10	11	11	11	11	11
$C=5$	3	6	10	14	14	14	14	14
$C=6$	3	6	10	15	17	17	17	17
$C=7$	3	6	10	15	20	20	20	20
$C=8$	3	6	10	15	21	23	23	23
$C=9$	3	6	10	15	21	26	26	26
$C=10$	3	6	10	15	21	28	29	29

(b) Model deficiencies

	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$	$K=8$	$K=9$	$K=10$
$C=2$	2	0	0	0	0	0	0	0
$C=3$	5	2	0	0	0	0	0	0
$C=4$	8	5	1	0	0	0	0	0
$C=5$	11	8	4	0	0	0	0	0
$C=6$	14	11	7	2	0	0	0	0
$C=7$	17	14	10	5	0	0	0	0
$C=8$	20	17	13	8	2	0	0	0
$C=9$	23	20	16	11	5	0	0	0
$C=10$	26	23	19	14	8	1	0	0

Table 4.2: Model ranks obtained from using the SN method for model $\{[\phi(h), p(h)]_C\}$ for different dimensions of K and C shown. To the left of the solid line, models are parameter–redundant and to the right of the solid line, models are not parameter–redundant. The dashed lines divides the models according to the rule of $K \geq 2C$.

(a) Model ranks based on the linear link. The same result is obtained using the logistic link function.

	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$	$K=8$	$K=9$	$K=10$
$C=2$	4	9	12	15	18	21	24	27
$C=3$	4	9	16	21	25	29	33	37
$C=4$	4	9	18	27	32	37	42	47
$C=5$	4	9	18	33	39	45	51	57
$C=6$	4	9	18	35	46	53	60	67
$C=7$	4	9	18	35	53	61	69	77
$C=8$	4	9	18	35	60	69	78	87
$C=9$	4	9	18	35	67	77	87	97
$C=10$	4	9	18	35	68	85	96	107

(b) Model deficiencies

	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$	$K=8$	$K=9$	$K=10$
$C=2$	3	1	1	1	1	1	1	1
$C=3$	6	5	2	1	1	1	1	1
$C=4$	9	9	5	1	1	1	1	1
$C=5$	12	13	10	1	1	1	1	1
$C=6$	15	17	15	5	1	1	1	1
$C=7$	18	21	20	11	1	1	1	1
$C=8$	21	25	25	17	1	1	1	1
$C=9$	24	29	30	23	1	1	1	1
$C=10$	27	33	35	29	7	1	1	1

Table 4.3: Model ranks obtained from using the SN method for model $\{\phi(t), p(t \times h_C)\}$ for different dimensions of K and C shown. To the left of the solid line, models are parameter–redundant due to both lack of data and the model structure, and to the right of the solid line, models are parameter–redundant due to the model structure only. The dashed lines divides the models according to the rule of $K \geq 2C$.

We note that the model has a deficiency ≥ 1 for all dimensions of K and C given. The model is always parameter redundant. The solid line divides the table. To the right of the line, the model is parameter redundant due to the model structure itself. To the left of the line, the model is parameter redundant due to both the limited information we have and the model structure itself.

That is, to the right of the solid line, the model is parameter–redundant due to its construction. No matter how many years of data we collect, the model is always parameter–redundant and we need to put extra constraints to estimate the parameters. To the left of the solid line, the model is parameter–redundant due to the fact there are not enough capture histories as well as the construction of the model.

We note that in Pledger *et al.* (2003), the authors have got most of the estimable parameters correct as given in Table 1 of their paper. The only differences are: for the model $\{[\phi(t+h), p(t \times h)]_C\}$, we find the same number of estimable parameters as $KC + K + C - 4$, but note there is a deficiency of 1; for the model $\{[\phi(t \times h), p(t + h)]_C\}$, we find the number of estimable parameters as $KC + C + K - 4$, but note there is also a deficiency of 1; Both differences might due to typo errors. This can occur as they give the total number of estimable parameters for the model $\{\phi(t + h_C), p(t)\}$ to be $2K_2C - 4$, whereas it should be $2K + 2C - 4$.

Finally, we note that the *number of independent estimable parameters* given in Table 1 by Pledger *et al.* (2003) are correct only if certain conditions are met. For example in Table 4.3, we can deduce that the number of estimable parameters is $KC + K - 3$. Pledger *et al.* (2003) have found the same result. But this is only valid to the right of the solid line in the table. What they have not found, are the values to the left of the solid lines. We will discuss how to find the exact rule and general conditions in Chapter 5.

We have found that with the exception of the models $\{[\phi(h), p(t+h)]_C\}$ and $\{[\phi(t+h), p(h)]_C\}$ among those CJS mixture models in Table 3.2, the same model ranks (and deficiencies) will be obtained from using the linear link function and using the logistic link function. Table 4.4 shows a different case, where using different link functions can result in different model structures, and therefore different parameter–redundancy results.

In Table 4.4, for the model $\{[\phi(h), p(t+h)]_C\}$, the ranks agree for $C = 2, 3$, but start to disagree from $C = 4$ and $K = 4$. Using the linear link function, the model has rank 10 while using the logistic link function, the model has rank 11 for $K = 4, C = 4$. For the model $\{[\phi(t+h), p(t)]_C\}$, model ranks start to disagree from $K = 4, C = 3$.

We observe that for the model $\{[\phi(h), p(t+h)]_C\}$, the model ranks are smaller using the linear link function than using the logistic link function whenever they disagree. For the model $\{[\phi(t+h), p(t)]_C\}$, the model ranks are larger using the linear link function than using the logistic link function whenever they disagree. This is because that there is no one–to–one relationship between the two link functions for both models. We will provide more details later in Example 5.2.1 on one–to–one transformation between the two link functions.

For ease of coding purposes, we have used the same link functions in the same model. That is, we have not looked at the case where one parameter uses the linear link function while the other uses the logistic link function. But it could be interesting to do so and investigate. Not that the aim is to find whether the original model (when the logistic link is used) is full rank. So in practise, we prefer to use the logistic link function for both parameters when probabilities are involved.

(a) Model $\{[\phi(h), p(t+h)]_C\}$: ranks based on the linear link

	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$	$K=8$	$K=9$	$K=10$
$C=2$	4	7	8	9	10	11	12	13
$C=3$	4	10	11	12	13	14	15	16
$C=4$	4	10	14	15	16	17	18	19
$C=5$	4	10	16	18	19	20	21	22
$C=6$	4	10	16	21	22	23	24	25
$C=7$	4	10	16	23	25	26	27	28
$C=8$	4	10	16	23	28	29	30	31
$C=9$	4	10	16	23	31	32	33	34
$C=10$	4	10	16	23	31	35	36	37

(b) Model $\{[\phi(h), p(t+h)]_C\}$: ranks based on the logistic link

	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$	$K=8$	$K=9$	$K=10$
$C=2$	4	7	8	9	10	11	12	13
$C=3$	4	10	11	12	13	14	15	16
$C=4$	4	11	14	15	16	17	18	19
$C=5$	4	11	17	18	19	20	21	22
$C=6$	4	11	19	21	22	23	24	25
$C=7$	4	11	19	24	25	26	27	28
$C=8$	4	11	19	27	28	29	30	31
$C=9$	4	11	19	29	31	32	33	34
$C=10$	4	11	19	29	34	35	36	37

(c) Model $\{[\phi(t+h), p(t)]_C\}$: ranks based on the linear link

	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$	$K=8$	$K=9$	$K=10$
$C=2$	4	7	8	9	10	11	12	13
$C=3$	4	10	11	12	13	14	15	16
$C=4$	4	10	14	15	16	17	18	19
$C=5$	4	10	17	18	19	20	21	22
$C=6$	4	10	17	21	22	23	24	25
$C=7$	4	10	17	24	25	26	27	28
$C=8$	4	10	17	25	28	29	30	31
$C=9$	4	10	17	25	31	32	33	34
$C=10$	4	10	17	25	34	35	36	37

(c) Model $\{[\phi(t+h), p(t)]_C\}$: ranks based on the logistic link

	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$	$K=8$	$K=9$	$K=10$
$C=2$	4	7	8	9	10	11	12	13
$C=3$	4	8	11	12	13	14	15	16
$C=4$	4	8	13	15	16	17	18	19
$C=5$	4	8	13	18	19	20	21	22
$C=6$	4	8	13	19	22	23	24	25
$C=7$	4	8	13	19	25	26	27	28
$C=8$	4	8	13	19	26	29	30	31
$C=9$	4	8	13	19	26	32	33	34
$C=10$	4	8	13	19	26	34	36	37

Table 4.4: Different model ranks obtained from using the linear link function and the logistic link function using the SN method for K and C shown.

4.4 Theorems Needed for Formal Proof

The parameter–redundancy results from using the SN method are only valid for values of K and C given in the tables, and cannot be used as a proof for a general rule to detect parameter–redundancy. To develop a general rule for the results of parameter–redundancy, we need simpler exhaustive summaries and that is the topic to be discussed in Chapter 5.

A general framework on using an exhaustive summary to determine features of parameter–redundancy is illustrated in Figure 4.5, adapted and edited from Cole *et al.* (2010). By extending K and/or C, we extend the exhaustive summaries too. Consider an exhaustive summary $\boldsymbol{\kappa}_1(\boldsymbol{\theta}_1)$ with derivative matrix $\boldsymbol{D}_1(\boldsymbol{\theta}_1)$. We extend the model to $\boldsymbol{\kappa}(\boldsymbol{\theta}') = [\boldsymbol{\kappa}_1(\boldsymbol{\theta}_1), \boldsymbol{\kappa}_2(\boldsymbol{\theta}')]^T$ where $\boldsymbol{\theta}' = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2]$, by introducing new parameters $\boldsymbol{\theta}_2$. The derivative matrix of the new model is then

$$\boldsymbol{D} = \begin{bmatrix} \dfrac{\partial \boldsymbol{\kappa}_1(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1} & \dfrac{\partial \boldsymbol{\kappa}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_1} \\ \dfrac{\partial \boldsymbol{\kappa}_1(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_2} & \dfrac{\partial \boldsymbol{\kappa}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{D}_1 & \boldsymbol{D}_{2,1} \\ 0 & \boldsymbol{D}_2 \end{bmatrix}.$$

Theorem 4.4.1 (Extension Theorem)**.** *If the matrix $\boldsymbol{D}_1$ is full rank and $\boldsymbol{D}_2$ is also full rank, then $\boldsymbol{D}$ is full rank. It follows that the extended model is full rank.*

The extension theorem was developed by Catchpole and Morgan (1997). What we have presented here is Theorem 3 in Cole *et al.* (2010) in a generalised form. Using the extension theorem we can extend the model in both K and C.

As have seen in Example 4.2.3, a PLUR decomposition can be used to check whether a matrix is full rank: we here give the following theorem.

Theorem 4.4.2. *For any full rank model with derivative matrix $\boldsymbol{D}$, the model is parameter redundant at $\boldsymbol{\theta}$ if and only if the determinant of U is 0 at $\boldsymbol{\theta}$,*

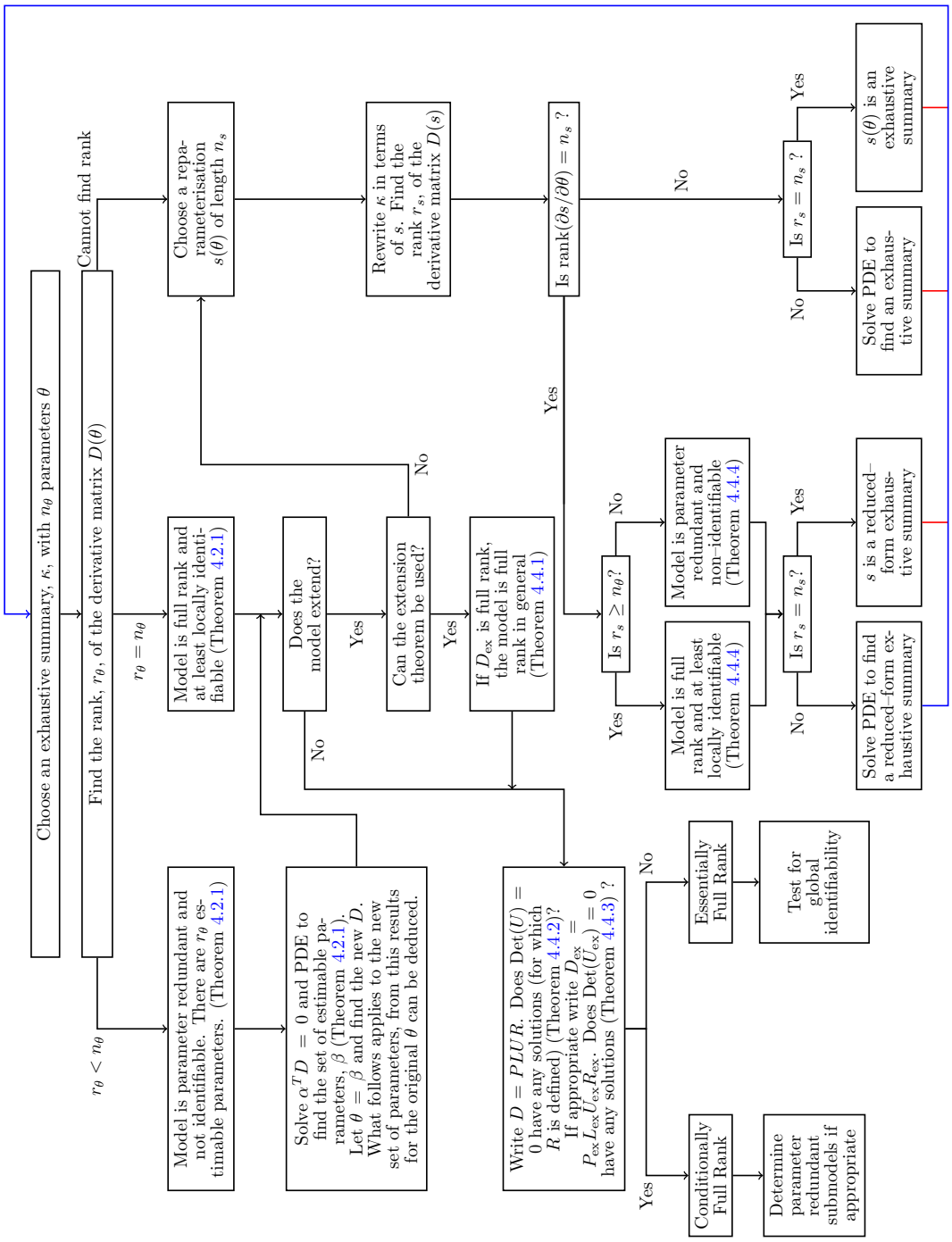

Figure 4.5: A general framework for detecting parameter–redundancy using exhaustive summaries. Adapted and edited from Cole *et al.* (2010).

where U is the upper triangular matrix from the PLUR decomposition of $\boldsymbol{D}$.

The above theorem is Theorem 4 in Cole *et al.* (2010).

If we write $\boldsymbol{D}_1 = P_1 L_1 U_1 R_1$ and $\boldsymbol{D}_2 = P_2 L_2 U_2 R_2$, then we have the following theorem.

Theorem 4.4.3. *If $Det(U_1) = 0$ or $Det(U_2) = 0$ at a point $\boldsymbol{\theta}$, then the extended model is parameter redundant at $\boldsymbol{\theta}$.*

The above theorem is Theorem 6 in Cole *et al.* (2010). This is useful when we extend the model using the extension theorem.

Definition 4.4.1. We define a *reduced–form exhaustive summary* to be a reparameterisation $\boldsymbol{s}(\boldsymbol{\theta})$ such $\boldsymbol{D}(\boldsymbol{s}) = \partial \boldsymbol{s}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ is full rank.

For example, the reparameterisation $\boldsymbol{s}(\boldsymbol{\theta}) = [\phi_1, p_2, \phi_2 p_3]^T$ in Example 4.1.1 is a reduced–form exhaustive summary.

Theorem 4.4.4 (Reparameterisation Theorem)**.** *If $r_{\boldsymbol{s}} = n_{\boldsymbol{s}}$, $\boldsymbol{s}(\boldsymbol{\theta})$ is a reduced–form exhaustive summary. If $r_{\boldsymbol{s}} < n_{\boldsymbol{s}}$, a reduced–form exhaustive summary can be found by solving $\boldsymbol{\alpha}(\boldsymbol{s})^T \boldsymbol{D}(f(\boldsymbol{s})) = 0$. If rank $(\partial \boldsymbol{s}/\partial \boldsymbol{\theta}) = n_{\boldsymbol{s}}$, there are $r_{\boldsymbol{s}}$ estimable parameters. If $r_{\boldsymbol{s}} = n_{\boldsymbol{\theta}}$, the model $M(\boldsymbol{\theta})$ is full rank. If $r_{\boldsymbol{s}} < n_{\boldsymbol{\theta}}$, the model $M(\boldsymbol{\theta})$ is parameter redundant.*

The above theorem is Theorem 8 in Cole *et al.* (2010). A reduced–form exhaustive summary is an exhaustive summary that can no longer be reduced in dimension. It is the smallest exhaustive summary and also provides the estimable parameters. We shall discuss more details on reduced–form exhaustive summaries in Section 5.3.

To summarise, for the open population mixture models, we will start by considering the smallest K and C.

Step 1. Start by finding a reparameterisation $\boldsymbol{s}(\boldsymbol{\theta})$.

Step 2. Show that $\boldsymbol{s}(\boldsymbol{\theta})$ is a (reduced–form) exhaustive summary using Theorem 4.4.4.

Step 3. We extend both K and C by 1, and show that $\boldsymbol{s}(\boldsymbol{\theta})$ is a (reduced–form) exhaustive summary using Theorem 4.4.1. It follows that $\boldsymbol{s}(\boldsymbol{\theta})$ is a (reduced–form) exhaustive summary.

This gives us a simpler exhaustive summary. We can then use $\boldsymbol{s}(\boldsymbol{\theta})$ to determine the parameter–redundancy result. We give examples of simpler exhaustive summaries in Chapter 5.

4.5 Conclusion

To conclude, parameter–redundancy arises when we cannot estimate all model parameters. For open population mixture models, the number of model parameters can increase with K and C, where K is the number of samples and C is the number of mixture components. Identifying the parameter–redundancy problem is typically not as easy as the example seen in Example 4.1.1.

Symbolic methods have been developed to detect such problems. Even with the aid of a computer package such as Maple® or Mathematica® , the model rank could be wrongly obtained without a further check, using a PLUR decomposition for example. The symbolic method relies on the ability to calculate the symbolic rank of a derivative matrix. This is not always possible and can fail due to the lack of computer memory and the complicated structure of the derivative matrix.

The hybrid symbolic–numerical (SN) method is developed so that we can determine model ranks more quickly and easily than using the symbolic method. This is because computer packages can calculate the rank of a matrix numer-

ically quicker than the rank of a symbolic matrix. Using the SN method, we can determine model ranks for any given K and C. We examined tables of model ranks (and deficiencies) to discover a pattern between K and C and the parameter–redundancy results.

Both the symbolic method and the SN method can determine which parameters are estimable but only the symbolic method can identify how the non–estimable parameters are combined together to give the estimable combinations.

As given in Holzmann *et al.* (2006), in a slightly different context, the rule of $K \geq 2C$ is suggested to ensure that models are identifiable for mixture models in closed populations. From the results using the SN method, we have seen that this rule can be applied for open population mixture models if only one of $\{\phi_{jc}\}$ or $\{p_{jc}\}$ is heterogeneous. General rules for these models are the topic of the next chapter and the theorems involved are listed in Section 4.4.

4.6 Code & List of files

In this section, we list all relevant files on the supplementary CD in Table 4.5. Details on using the Maple® package `MixtureModelforCaptureRecapture.mla` are given in Appendix A and details on using the Mathematica® package `MixtureModelforCaptureRecapture.m` are given in Appendix B.

For Example 4.2.3, we review the code that was originally used in Gimenez *et al.* (2003) and tried the same code in Maple® 6. We verified that the wrong rank 9 was given. After using `map(expand,matrixA)`, we were able to get a correct rank: see file `Maple_V6.pdf`. This is still the case as shown in the example: see file `Maple_V18.pdf`. For comparison, the same matrix rank was calculated in Mathematica® with a correct rank 8: see file `Mathematica_MatrixRank.pdf`.

File Name	Description
`Gimenez et al. (2003)`	The folder contains all source files from Gimenez *et al.* (2003)
`Maple_V6.pdf`	Rank of matrix $A(\theta)$ from using Maple® 6.
`Maple_V18.pdf`	Rank of matrix $A(\theta)$ from using Maple® 18.
`Mathematica_MatrixRank.pdf`	Rank of matrix $A(\theta)$ from using Mathematica® 10.
`Rank.mw`	Maple® worksheet for calculating the rank of matrix $A(\theta)$.
`MatrixRank.nb`	Mathematica® notebook for calculating the rank of matrix $A(\theta)$.
`Chapter_4.pdf`	Examples in this chapter using Mathematica® .
`model_ranks_linear.pdf`	Results of parameter–redundancy for CJS mixture models using the SN method when the linear link function is used.
`model_ranks_logit.pdf`	Results of parameter–redundancy for CJS mixture models using the SN method when the logistic link function is used.
`model_ranks_linear_JS.pdf`	Results of parameter–redundancy for JS mixture models using the SN method when the linear link function is used.
`model_ranks_logit_JS.pdf`	Results of parameter–redundancy for JS mixture models using the SN method when the logistic link function is used.

Table 4.5: List of files on Chapter 4

To utilize all 32 available cores on the school server `Emmy`, Mathematica® 10 was used. All notebooks (the Mathematica® file) are timed for completing the whole job. See individual files for more details. We note that it takes 769.18 minutes to finish for all 25 JS mixture models using the SN method when the logistic link function is used. This is nearly 13 hours. However we note that there were other jobs running on the server at the same time.

Chapter 5

Simpler Exhaustive Summaries

In Chapter 3 we have seen that open population mixture models can have many parameters and in Chapter 4 we showed that we cannot always estimate all model parameters and that this is the problem of parameter redundancy. We have seen how we can detect parameter redundancy using both the symbolic method and the hybrid symbolic–numerical (SN) method. Using the SN method, we can efficiently determine the parameter redundancy result for any given K and C, however it cannot be used to find general results.

In Chapter 4, we have also seen that the standard symbolic method fails in calculating the rank of the symbolic derivative matrix for large values of K and C. The main reason is that the exhaustive summaries are structurally too complex. To be able to use the symbolic method, we need to develop simpler exhaustive summaries. The simpler exhaustive summaries are exhaustive summaries which carry the same information as the original exhaustive summaries, but are either simpler in structure, or have fewer terms to consider. In this chapter, we provide some examples of simpler exhaustive summaries for various models discussed in Chapter 3.

In Section 5.1, we will start with examples using the *linear link function* and consider a range of open capture–recapture mixture models, starting with

relatively simple cases and then progressing to more complex ones. Models using the linear link function are structurally simpler and in many cases, where there is a one–to–one transformation, the parameter redundancy results are identical to the same models using a logistic link function; see Section 5.2.1.

Using examples from the CJS mixture models, we first describe how to find a reparameterisation that is *useful* in developing a simpler exhaustive summary in Section 5.1.1 using the model $\{\phi(t+h_C), p(\cdot)\}$. We then show the full proof of a simpler exhaustive summary for the model the $\{\phi(t), p(t+h_C)\}$ and all of its sub–models in Section 5.1.2. We show how the simpler exhaustive summary can be used to determine the parameter redundancy results for the model $\{\phi(t), p(t+h_C)\}$ and all of its sub–models in Section 5.1.3. We show the simpler exhaustive summary for a more complicated example where heterogeneity exists in both the survival probabilities and the recapture probabilities in Section 5.1.4 for the model $\{[\phi(h), p(h)]_C\}$. We further show what happens to the exhaustive summary when we introduce a time component in the model $\{[\phi(h), p(t+h)]_C\}$ in Section 5.1.5.

An example of a simpler exhaustive summary for the capture recovery model $\{\phi(h_C), \lambda(t)\}$ is given in Section 5.1.6. We show how the simpler exhaustive summary can be extended when we introduce covariates into the model by looking at the model $\{\phi(h_C+a), \lambda(t)\}$ in Section 5.1.7.

As the models get more complicated, the $H(x, y, z)$ function will be introduced as a convenient short–hand in Section 5.1.8 and we show how it is used in Section 5.1.9, through an example from the capture–recapture–recovery model $\{\phi(h_C), p(h_C), \lambda(h_C)\}$.

Moving on to Section 5.2, we will provide some examples of simpler exhaustive summaries for models using the *logistic link function*, which are more complicated than using the linear link function. In Section 5.2.1, we show the link between the linear link function and the logistic link function. We

demonstrate that if there is a one–to–one transformation, the parameter redundancy result will be the same. In Section 5.2.2, we prove the simpler exhaustive summary for the model $\{\phi(\cdot), p(t+h_C)\}$, for which there is no one–to–one transformation between the linear link function and the logistic link function. We further explore the more complicated model $\{\phi(t), p(t+h_C)\}$ in Section 5.2.3.

Section 5.3 discusses how the simpler exhaustive summaries can be used to help us determine parameter redundancy results and Section 5.4 concludes the findings in this chapter.

5.1 Using the Linear Link Function

In this section, we will provide some examples of simpler exhaustive summaries from using the linear link function. Examples will range from relatively simple open population mixture models, to more complex models, including models with covariates.

5.1.1 How to find a reparameterisation?

In this section, we will provide an example to show how to find a reparameterisation that can be used to find a simpler exhaustive summary for the model $\{\phi(t+h_C), p(\cdot)\}$. The model is constant in recapture probabilities and time–varying and heterogeneous in survival probabilities. We use the linear link function to motivate the example as it is simpler than the logistic link function. When the linear link function is used, the heterogeneous survival probabilities are as follows

$$\phi_{jc} = \mu_\phi + \tau_{\phi j} + \eta_{\phi c},$$

where μ_ϕ is the baseline from group 1, $\{\tau_{\phi j}\}$ are the time–varying components and $\{\eta_{\phi c}\}$ are the heterogeneous components.

Using Theorem 4.4.4 we find the estimable parameters by solving the system of partial differential equations (PDEs) as have seen in Example 4.2.1. Using Maple® to solve the PDE system we found that the estimable parameters for $K = 3, C = 2$ are given by

$$\begin{cases} p \\ \tau_{\phi 2} \\ \eta_{\phi 2}\sqrt{w_1(w_1-1)} \\ \mu_\phi + \eta_{\phi 2} - w_1\eta_{\phi 2} \end{cases}.$$

This is a reparameterisation. We note that the $\eta_{\phi 2}\sqrt{w_1(w_1-1)}$ is going to produce a complex number since the expression in the square root will be negative for any sensible values of w_1. But this is a perfectly valid symbolic solutions for the PDE system. This raises a question of whether we can always make sense of the reparameterisation.

The single set shown above does not help us much in determining a general reparameterisation as we do not observe a general pattern. We can find the estimable parameters using Maple® for alternative values of K and C, for example, for $K = 3, C = 3$, the estimable parameters are

$$\begin{cases} p \\ \tau_{\phi 2} \\ (-1+w_1+w_2)\eta_{\phi 3} - w_2\eta_{\phi 2} - \mu_\phi \\ (1-w_1-w_2)\eta_{\phi 3}^2 - 2\mu_\phi(-1+w_1+w_2)\eta_{\phi 3} + 2\eta_{\phi 2}\mu_\phi w_2 + \eta_{\phi 2}^2 w_2 + \mu_\phi^2 \end{cases},$$

which can be rearranged to obtain

$$\begin{cases} p \\ \tau_{\phi 2} \\ -w_3\eta_{\phi 3} - w_2\eta_{\phi 2} - \mu_\phi \\ w_3(\mu_\phi + \eta_{\phi 3})^2 + w_2(\mu_\phi + \eta_{\phi 2})^2 + w_1\mu_\phi^2 \end{cases}.$$

For $K = 3, C = 4$ the estimable parameters can be rearranged to give

$$\begin{cases} p \\ \tau_{\phi 2} \\ -w_4\eta_{\phi 4} - w_3\eta_{\phi 3} - w_2\eta_{\phi 2} - \mu_\phi \\ w_4(\mu_\phi + \eta_{\phi 4})^2 + w_3(\mu_\phi + \eta_{\phi 3})^2 + w_2(\mu_\phi + \eta_{\phi 2})^2 + w_1\mu_\phi^2 \end{cases}.$$

We can now observe a general pattern when C increases, except the case when $C = 2$. But we will show later that the general pattern will work when $C = 2$ (and for $C = 1$) too. It is just a *different* reparemeterisation than the one found by solving the PDE system. So now we explore the pattern by extending to $K = 4$. The model is full rank when $K = 4, C = 2$, so that all parameters can be estimated in this case. We find that Maple® fails to solve the PDE system for $K = 4, C = 3$ and returns with a warning:

Warning: Incomplete separation.

We continue to try for $K = 4, C = 4$ and obtain the estimable parameters

$$\begin{cases} p \\ \tau_{\phi 2} \\ \tau_{\phi 3} \\ -w_4\eta_{\phi 4} - w_3\eta_{\phi 3} - w_2\eta_{\phi 2} - \mu_\phi \\ w_4(\mu_\phi + \eta_{\phi 4})^2 + w_3(\mu_\phi + \eta_{\phi 3})^2 + w_2(\mu_\phi + \eta_{\phi 2})^2 + w_1\mu_\phi^2 \\ w_4(\mu_\phi + \eta_{\phi 4})^3 + w_3(\mu_\phi + \eta_{\phi 3})^3 + w_2(\mu_\phi + \eta_{\phi 2})^3 + w_1\mu_\phi^3 \end{cases},$$

after rearranging. We can try a few more sets of values for K and C and investigate the estimable parameters. By rearranging the estimable parameters and simplifying terms, we can deduce a pattern and find that a possible reparameterisation for the model $\{\phi(t + h_C), p(\cdot)\}$ when the linear link function is used is given by

$$\begin{bmatrix} p \\ \tau_{\phi 2} \\ \vdots \\ \tau_{\phi K-1} \\ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c}) \\ \vdots \\ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^{K-1} \end{bmatrix}. \tag{5.1}$$

To verify this, we consider the heterogeneous survival probabilities $\{\phi_{jc}\}$

at each time occasion j, which are determined by

$$\begin{aligned}
\phi_j &= \sum_{c=1}^{C} w_c \phi_{jc} && \text{(5.2a)}\\
&= \sum_{c=1}^{C} w_c \left(\mu_\phi + \tau_{\phi j} + \eta_{\phi c}\right) && \text{(5.2b)}\\
&= \sum_{c=1}^{C} \left\{ w_c \tau_{\phi j} + w_c \left(\mu_\phi + \eta_{\phi c}\right) \right\} && \text{(5.2c)}\\
&= \tau_{\phi j} + \sum_{c=1}^{C} w_c \left(\mu_\phi + \eta_{\phi c}\right), && \text{(5.2d)}
\end{aligned}$$

as $\sum_{c=1}^{C} w_c = 1$. Recall that we set $w_C = 1 - \sum_{c=1}^{C-1} w_c$ and $\eta_{\phi 1} = 0$ as constraints in this model.

We will now show that the reparameterisation given in (5.1) is a simpler exhaustive summary for the model $\{\phi(t + h_C), p(\cdot)\}$ using the extension theorem, Theorem 4.4.1. We start by considering the original exhaustive summary for $K = 3, C = 1$: we have

$$\boldsymbol{\kappa}(\boldsymbol{\theta}) = \begin{bmatrix} 1 \\ 1 - \mu_\phi - \tau_{\phi 2} + (\mu_\phi + \tau_{\phi 2})\overline{p} \\ (\mu_\phi + \tau_{\phi 2})p \\ 1 - \mu_\phi + \mu_\phi \bar{p}(1 - \mu_\phi - \tau_{\phi 2}) + \mu_\phi \overline{p}^2(\mu_\phi + \tau_{\phi 2}) \\ \mu_\phi \bar{p}(\mu_\phi + \tau_{\phi 2})p \\ \mu_\phi p(1 - \mu_\phi - \tau_{\phi 2}) + \mu_\phi \overline{p}(\mu_\phi + \tau_{\phi 2})p \\ \mu_\phi p^2(\mu_\phi + \tau_{\phi 2}) \end{bmatrix}, \qquad (5.3)$$

where as previously, $\overline{\theta} = 1 - \theta$ for any parameter θ. This reparameterisation

is given by

$$\boldsymbol{s}(\boldsymbol{\theta}) = \begin{bmatrix} p \\ \tau_{\phi 2} \\ \mu_\phi \\ \mu_\phi^2 \end{bmatrix}. \tag{5.4}$$

The reparameterisation has more terms than necessary, but we will use this as an illustration. Now we rewrite the original exhaustive summary $\boldsymbol{\kappa}(\boldsymbol{\theta})$ using $\boldsymbol{s}(\boldsymbol{\theta})$ to obtain

$$\boldsymbol{\kappa}(\boldsymbol{s}) = \begin{bmatrix} 1 \\ -(s_2+s_3)s_1+1 \\ (s_2+s_3)s_1 \\ (s_2s_3+s_4)s_1^2+(-s_2s_3-s_3-s_4)s_1+1 \\ (-s_2s_2-s_4)s_1^2+(s_2s_3+s_4)s_1 \\ (-s_2s_2-s_4)s_1^2+s_3s_1 \\ (s_2s_3+s_4)s_1^2 \end{bmatrix}. \tag{5.5}$$

Then we differentiate the $\boldsymbol{\kappa}(\boldsymbol{s})$ with respect to $\boldsymbol{s}$, to obtain the derivative matrix

$$\boldsymbol{D}(\boldsymbol{s}) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ -s_2-s_3 & -s_1 & -s_1 & 0 \\ s_2+s_3 & s_1 & s_1 & 0 \\ \vdots & & & \vdots \\ 2(s_2s_3+s_4)s_1 & s_3s_1^2 & s_2s_1^2 & s_1^2 \end{bmatrix}. \tag{5.6}$$

We find using Maple® that the derivative matrix $\boldsymbol{D}(\boldsymbol{s})$ has rank $r_{\boldsymbol{s}} = 4$. The reparameterisation $\boldsymbol{s}(\boldsymbol{\theta})$ has dimension $n_{\boldsymbol{s}} = 4$. So the reparameterisation $\boldsymbol{s}(\boldsymbol{\theta})$ is a simpler exhaustive summary for $K = 3, C = 1$ for the model $\{\phi(t+h_C), p(\cdot)\}$, using Theorem 4.4.4.

We now keep $K = 3$ and extend to $C = 2$. The original exhaustive sum-

mary is given by

$$\boldsymbol{\kappa}(\boldsymbol{\theta}) = \begin{bmatrix} 1 \\ \vdots \\ w_1(\mu_\phi + \tau_{\phi 2})p + (1 - w_1)(\mu_\phi + \tau_{\phi 2} + \eta_{\phi 2})p \\ \vdots \\ w_1\mu_\phi p^2(\mu_\phi + \tau_{\phi 2}) + (1 - w_1)\mu_\phi p^2(\mu_\phi + \tau_{\phi 2} + \eta_{\phi 2}) \end{bmatrix}. \tag{5.7}$$

The reparameterisation is

$$\boldsymbol{s}(\boldsymbol{\theta}) = \begin{bmatrix} p \\ \tau_{\phi 2} \\ w_1\mu_\phi + (1 - w_1)(\mu_\phi + \eta_{\phi 2}) \\ w_1\mu_\phi^2 + (1 - w_1)(\mu_\phi + \eta_{\phi 2})^2 \end{bmatrix}. \tag{5.8}$$

Rewriting $\boldsymbol{\kappa}(\boldsymbol{\theta})$ using $\boldsymbol{s}(\boldsymbol{\theta})$ we obtain

$$\boldsymbol{\kappa}(\boldsymbol{s}) = \begin{bmatrix} 1 \\ -(s_2 + s_3)s_1 + 1 \\ (s_2 + s_3)s_1 \\ (s_2s_3 + s_4)s_1^2 + (-s_2s_3 - s_3 - s_4)s_1 + 1 \\ (-s_2s_2 - s_4)s_1^2 + (s_2s_3 + s_4)s_1 \\ (-s_2s_2 - s_4)s_1^2 + s_3s_1 \\ (s_2s_3 + s_4)s_1^2 \end{bmatrix}. \tag{5.9}$$

We note that this is identical to the previous $\boldsymbol{\kappa}(\boldsymbol{s})$ for $K = 3, C = 1$. It follows that $\boldsymbol{D}(\boldsymbol{s})$ is full rank with $r_{\boldsymbol{s}} = 4$. Since $r_{\boldsymbol{s}} = n_{\boldsymbol{s}} = 4$, the reparameterisation $\boldsymbol{s}(\boldsymbol{\theta})$ is a simpler exhaustive summary for $K = 3, C = 2$ for the model $\{\phi(t + h_C), p(\cdot)\}$, using Theorem 4.4.4. We conclude that the reparameterisation in (5.1) is a simpler exhaustive summary for $K = 3$ and $C \geq 1$ using Theorem 4.4.1.

We will now show that we can extend the simpler exhaustive summary for K. We keep $C = 1$ and extend to $K = 4$. The original exhaustive summary is

$$\boldsymbol{\kappa}(\boldsymbol{\theta}) = \begin{bmatrix} 1 \\ \vdots \\ (\mu_\phi + \tau_{\phi 3})p \\ \vdots \\ \mu_\phi p^3(\mu_\phi + \tau_{\phi 2})(\mu_\phi + \tau_{\phi 3}) \end{bmatrix}. \tag{5.10}$$

The reparameterisation is given by

$$\boldsymbol{s}(\boldsymbol{\theta}) = \begin{bmatrix} p \\ \tau_{\phi 2} \\ \tau_{\phi 3} \\ \mu_\phi \\ \mu_\phi^2 \\ \mu_\phi^3 \end{bmatrix}. \tag{5.11}$$

The reparameterisation has more terms than necessary by construction, but we will use this as an illustration. Now we rewrite the original exhaustive summary $\boldsymbol{\kappa}(\boldsymbol{\theta})$ using $\boldsymbol{s}(\boldsymbol{\theta})$ to obtain

$$\boldsymbol{\kappa}(\boldsymbol{s}) = \begin{bmatrix} 1 \\ -(s_3 + s_4)s_1 + 1 \\ (s_3 + s_4)s_1 \\ \vdots \\ (s_2 s_3 s_4 + s_2 s_5 + s_3 s_5 + s_6)s_1^3 \end{bmatrix}. \tag{5.12}$$

We differentiate $\boldsymbol{\kappa}(\boldsymbol{s})$ with respect to the elements of $\boldsymbol{s}$ to obtain the derivative

matrix

$$\boldsymbol{D}(\boldsymbol{s}) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ -(s_3+s_4) & 0 & -s_1 & -s_1 & 0 & 0 \\ \vdots & \cdots & \cdots & \cdots & \cdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & s_1^3 \end{bmatrix}. \tag{5.13}$$

We found using Maple® that the derivative matrix $\boldsymbol{D}(\boldsymbol{s})$ has rank $r_{\boldsymbol{s}} = 6$. Since the dimension of $\boldsymbol{s}(\boldsymbol{\theta})$ is $n_{\boldsymbol{s}} = 6$, so that $r_{\boldsymbol{s}} = n_{\boldsymbol{s}}$, we deduce that $\boldsymbol{s}(\boldsymbol{\theta})$ is a simpler exhaustive summary for $K = 4, C = 1$. Using the extension theorem, Theorem 4.4.1, the reparameterisation in Equation (5.1) is a simpler exhaustive summary for the model $\{\phi(t+h_C), p(\cdot)\}$ for $K \geq 3, C \geq 1$.

5.1.2 Model: $\{\phi(t), p(t+h_C)\}$

In this model, the survival probabilities are time–varying, denoted by $\{\phi_j\}$. The capture probabilities are time–varying and heterogeneous, denoted by

$$p_{jc} = \mu_p + \tau_{pc} + \eta_{pc},$$

where a linear link function has been used. Here, μ_p is the baseline from group 1, τ_{pc} is the time component and η_{pc} is the heterogeneous component, both for class c.

We first examine parameter redundancy using the hybrid symbolic–numerical (SN) method described in Section 4.2.4. We obtain the table of model ranks and deficiencies as given in Table 5.1.

We note in the table that from $K = 4$, the model ranks $8, 11, 14, \ldots, 26$ form an arithmetic sequence with a difference of 3. So we will show in the example below, that from $K \geq 4$ and $C \geq 2$, a simpler exhaustive summary can be derived using the extension theorem given in Theorem 4.4.1.

(a) Model ranks based on the linear link (same result using the logistic link)

	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$	$K=8$	$K=9$	$K=10$
$C=2$	4	8	10	12	14	16	18	20
$C=3$	4	8	11	14	16	18	20	22
$C=4$	4	8	11	14	17	20	22	24
$C=5$	4	8	11	14	17	20	23	26
$C=6$	4	8	11	14	17	20	23	26
$C=7$	4	8	11	14	17	20	23	26
$C=8$	4	8	11	14	17	20	23	26
$C=9$	4	8	11	14	17	20	23	26
$C=10$	4	8	11	14	17	20	23	26

(b) Model deficiencies

	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$	$K=8$	$K=9$	$K=10$
$C=2$	2	0	0	0	0	0	0	0
$C=3$	4	2	1	0	0	0	0	0
$C=4$	6	4	3	2	1	0	0	0
$C=5$	8	6	5	4	3	2	1	0
$C=6$	10	8	7	6	5	4	3	2
$C=7$	12	10	9	8	7	6	5	4
$C=8$	14	12	11	10	9	8	7	6
$C=9$	16	14	13	12	11	10	9	8
$C=10$	18	16	15	14	13	12	11	10

Table 5.1: Model ranks obtained from using the SN method for model $\{\phi(t), p(t+h_C)\}$ for different dimensions of K and C shown. Models above the dashed line satisfying the inequality $K \geq 2C$ and models below the dashed line satisfying the inequality $K < 2C$.

Example 5.1.1. A possible reparameterisation for this model is

$$\boldsymbol{s}(\boldsymbol{\theta}) = \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_{K-1} \\ \tau_{p3} \\ \vdots \\ \tau_{pK} \\ \sum_{c=1}^{C} w_c(\mu_p + \eta_{pc}) \\ \vdots \\ \sum_{c=1}^{C} w_c(\mu_p + \eta_{pc})^{K-1} \end{bmatrix}. \tag{5.14}$$

We will show that $\boldsymbol{s}(\boldsymbol{\theta})$ is a simpler exhaustive summary for $K = 4, C = 2$ and can be extended to $K = 4, C = 3$ and $K = 5, C = 2$. Hence using Theorem 4.4.1 (the extension theorem), $\boldsymbol{s}(\boldsymbol{\theta})$ is a simpler exhaustive summary for this model for $K \geq 4$ and $C \geq 2$.

Let us start with $K = 4, C = 2$. In this case, the reparameterisation is given by

$$\boldsymbol{s}(\boldsymbol{\theta}) = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \\ s_7 \\ s_8 \end{bmatrix} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \tau_{p_3} \\ \tau_{p_4} \\ w_1\mu_p + (1-w_1)(\mu_p + \eta_{p2}) \\ w_1\mu_p^2 + (1-w_1)(\mu_p + \eta_{p2})^2 \\ w_1\mu_p^3 + (1-w_1)(\mu_p + \eta_{p2})^3 \end{bmatrix}. \tag{5.15}$$

The original exhaustive summary $\boldsymbol{\kappa}(\boldsymbol{\theta})$ is too large to display on paper, so we will omit it here. However it is available in the relevant computer supple-

mentary files as a Maple® worksheet. We can rewrite the original exhaustive summary $\boldsymbol{\kappa}(\boldsymbol{\theta})$ in terms of $\boldsymbol{s}(\boldsymbol{\theta})$ to obtain $\boldsymbol{\kappa}(\boldsymbol{s})$:

$$\begin{bmatrix} 1 \\ -s_3s_5 - s_3s_6 + 1 \\ s_3s_5 + s_3s_6 \\ s_2s_3s_4s_5 + s_2s_3s_4s_6 + s_2s_3s_5s_6 - s_2s_3s_5 - s_2s_3s_6 + s_2s_3s_7 - s_2s_4 - s_2s_6 + 1 \\ -s_2s_3s_4s_5 - s_2s_3s_4s_6 - s_2s_3s_5s_6 + s_2s_3s_5 + s_2s_3s_6 - s_2s_3s_7 \\ -s_2s_3s_4s_5 - s_2s_3s_4s_6 - s_2s_3s_5s_6 - s_2s_3s_7 + s_2s_4 + s_2s_6 \\ s_2s_3s_4s_5 + s_2s_3s_4s_6 + s_2s_3s_5s_6 + s_2s_3s_7 \\ \vdots \\ \vdots \\ (-s_2s_3s_4s_5s_6 - s_2s_3s_4s_7 + s_2s_3s_5s_6 - s_2s_3s_5s_7 + s_2s_3s_7 - s_2s_3s_8)s_1 \\ (-s_2s_3s_4s_5s_6 - s_2s_3s_4s_7 - s_2s_3s_5s_7 - s_2s_3s_8 + s_2s_4s_6 + s_2s_7)s_1 \\ (s_2s_3s_4s_5s_6 + s_2s_3s_4s_7 + s_2s_3s_5s_7 + s_2s_3s_8)s_1 \end{bmatrix}. \quad (5.16)$$

We calculate the derivative matrix $\boldsymbol{D}(\boldsymbol{s})$ and find its rank is $r_{\boldsymbol{s}} = 8$. Since $n_{\boldsymbol{s}} = 8$, we have that $r_{\boldsymbol{s}} = n_{\boldsymbol{s}}$. We conclude that $\boldsymbol{s}(\boldsymbol{\theta})$ in Equation (5.15) is an exhaustive summary for $K = 4, C = 2$ for the model $\{\phi(\cdot), p(t + h_C)\}$ using Theorem 4.4.4.

Since the model is full rank in terms of the reparameterisation $\boldsymbol{s}(\boldsymbol{\theta})$, we can also check for its sub–models using Theorem 4.4.2. We now do a PLUR (Corless and Jeffrey, 1997) decomposition of the derivative matrix $\boldsymbol{D}(\boldsymbol{s})$ using

Maple® to obtain

$$P = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}$$

and

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

We omit the matrices L and U due to the fact that they are too complicated and large to display on paper, see supplementary Maple® worksheet for full versions.

Of interest here is that the determinant of the matrix U is given by

$$-(s_6^4 + s_5 s_6^2 - 2s_6^2 s_7 - s_5 s_7 - s_6 s_7 + s_7^2) s_1^3 s_2^5 s_3^4 (s_5 + s_6). \tag{5.17}$$

We need to check that the conditions to obtain the sub–models do not result in the expression above being 0. For example, the sub–model $\{\phi(\cdot), p(t + h_C)\}$ is equivalent to having $s_1 = s_2 = s_3$. This condition does not make the determinant of the matrix U equal to 0. Therefore, the sub–model $\{\phi(\cdot), p(t + h_C)\}$ in terms of $\boldsymbol{s}(\theta)$ is also full rank using Theorem 4.4.2. It follows that $\boldsymbol{s}(\theta)$ is a

simpler exhaustive summary for the model $\{\phi(\cdot), p(t + h_C)\}$ for $K = 4, C = 2$ using Theorem 4.4.4.

We check all the sub–models using the following conditions:

- $\{\phi(\cdot), p(\cdot)\} : s_1 = s_2 = s_3, s_4 = s_5 = 0, s_6 = s_7 = s_8$
- $\{\phi(\cdot), p(t)\} : s_1 = s_2 = s_3, s_6 = s_7 = s_8$
- $\{\phi(\cdot), p(h_C)\} : s_1 = s_2 = s_3, s_4 = s_5 = 0$
- $\{\phi(\cdot), p(t + h_C)\} : s_1 = s_2 = s_3$
- $\{\phi(t), p(\cdot)\} : s_4 = s_5 = 0, s_4 = s_5 = 0, s_6 = s_7 = s_8$
- $\{\phi(t), p(t)\} : s_6 = s_7 = s_8$
- $\{\phi(t), p(h_C)\} : s_4 = s_5 = 0$

None of the conditions makes the determinant of U equal to 0. Therefore, all sub–models of the model $\{\phi(t), p(t + h_C)\}$ in terms of the reparameterisation $\boldsymbol{s}(\boldsymbol{\theta})$ in Equation (5.15) is full rank. It follows that $\boldsymbol{s}(\boldsymbol{\theta})$ in Equation (5.15) is an exhaustive summary for $K = 4, C = 2$ for all sub–models of $\{\phi(t), p(t+h_C)\}$ using Theorem 4.4.4.

We now extend the model to $C = 3$, keeping $K = 4$. The reparameterisa-

tion is

$$\boldsymbol{s}(\boldsymbol{\theta}) = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \\ s_7 \\ s_8 \end{bmatrix} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \tau_{p_3} \\ \tau_{p_4} \\ w_1\mu_p + w_2(\mu_p + \eta_{p2}) + (1 - w_1 - w_2)(\mu_p + \eta_{p3}) \\ w_1\mu_p^2 + w_2(\mu_p + \eta_{p2})^2 + (1 - w_1 - w_2)(\mu_p + \eta_{p3})^2 \\ w_1\mu_p^3 + w_2(\mu_p + \eta_{p2})^3 + (1 - w_1 - w_2)(\mu_p + \eta_{p3})^3 \end{bmatrix}. \quad (5.18)$$

We note that $\boldsymbol{\kappa}(\boldsymbol{s})$ is identical to the one in Equation (5.16). This is due to the fact that the new parameters $\{w_2, \eta_{p3}\}$ are absorbed in the $\{s_6, s_7, s_8\}$ by the construction of the reparameterisation. It follows that the derivative matrix $\boldsymbol{D}(\boldsymbol{s})$ is also of rank $r_{\boldsymbol{s}} = 8$. Therefore $\boldsymbol{s}(\boldsymbol{\theta})$ in Equation (5.18) is an exhaustive summary for $K = 4, C = 3$ for the model $\{\phi(t), p(t+h_C)\}$ using Theorem 4.4.4.

To check its sub–models, we need the PLUR decomposition of $\boldsymbol{D}(\boldsymbol{s})$. Since $\boldsymbol{\kappa}(\boldsymbol{s})$ is identical to the one in Equation (5.16), it follows that $\boldsymbol{D}(\boldsymbol{s})$ is also identical to the derivative matrix we obtained from $K = 4, C = 2$. Hence, we can conclude that $\boldsymbol{s}(\boldsymbol{\theta})$ in Equation (5.18) is an exhaustive summary for $K = 4, C = 3$ for all sub–models of $\{\phi(t), p(t + h_C)\}$ using Theorem 4.4.2.

We also extend the model to $K = 5$, keeping $C = 2$. The reparameterisa-

tion is given by

$$\boldsymbol{s}(\boldsymbol{\theta}) = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \\ s_7 \\ s_8 \\ s_9 \\ s_{10} \\ s_{11} \end{bmatrix} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 \\ \tau_{p_3} \\ \tau_{p_4} \\ \tau_{p_5} \\ w_1\mu_p + (1-w_1)(\mu_p + \eta_{p2}) \\ w_1\mu_p^2 + (1-w_1)(\mu_p + \eta_{p2})^2 \\ w_1\mu_p^3 + (1-w_1)(\mu_p + \eta_{p2})^3 \\ w_1\mu_p^4 + (1-w_1)(\mu_p + \eta_{p2})^4 \end{bmatrix}. \tag{5.19}$$

The new exhaustive summary $\boldsymbol{\kappa}(\boldsymbol{s})$ is given by

$$\begin{bmatrix} 1 \\ -s_4s_7 - s_4s_8 + 1 \\ s_4s_7 + s_4s_8 \\ s_3s_4s_6s_7 + s_3s_4s_6s_8 + s_3s_4s_7s_8 - s_3s_4s_7 - s_3s_4s_8 + s_3s_4s_9 - s_3s_6 - s_3s_8 + 1 \\ -s_3s_4s_6s_7 - s_3s_4s_6s_8 - s_3s_4s_7s_8 + s_3s_4s_7 + s_3s_4s_8 - s_3s_4s_9 \\ -s_3s_4s_6s_7 - s_3s_4s_6s_8 - s_3s_4s_7s_8 - s_3s_4s_9 + s_3s_6 + s_3s_8 \\ s_3s_4s_6s_7 + s_3s_4s_6s_8 + s_3s_4s_7s_8 + s_3s_4s_9 \\ \vdots \\ \vdots \end{bmatrix}. \tag{5.20}$$

We observe that the first seven terms are of similar structure as $\boldsymbol{\kappa}(\boldsymbol{s})$ in

Equation (5.16), subject to relabelling. We therefore relabel as

$$\boldsymbol{s}(\boldsymbol{\theta}) = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \\ s_7 \\ s_8 \\ s_9 \\ s_{10} \\ s_{11} \end{bmatrix} = \begin{bmatrix} \phi_2 \\ \phi_3 \\ \phi_4 \\ \tau_{p_4} \\ \tau_{p_5} \\ w_1\mu_p + (1-w_1)(\mu_p + \eta_{p2}) \\ w_1\mu_p^2 + (1-w_1)(\mu_p + \eta_{p2})^2 \\ w_1\mu_p^3 + (1-w_1)(\mu_p + \eta_{p2})^3 \\ \phi_1 \\ \tau_{p_3} \\ w_1\mu_p^4 + (1-w_1)(\mu_p + \eta_{p2})^4 \end{bmatrix}. \tag{5.21}$$

We obtain the updated $\boldsymbol{\kappa}(\boldsymbol{s})$ as

$$\begin{bmatrix} 1 \\ -s_3s_5 - s_3s_6 + 1 \\ s_3s_5 + s_3s_6 \\ s_2s_3s_4s_5 + s_2s_3s_4s_6 + s_2s_3s_5s_6 - s_2s_3s_5 - s_2s_3s_6 + s_2s_3s_7 - s_2s_4 - s_2s_6 + 1 \\ -s_2s_3s_4s_5 - s_2s_3s_4s_6 - s_2s_3s_5s_6 + s_2s_3s_5 + s_2s_3s_6 - s_2s_3s_7 \\ -s_2s_3s_4s_5 - s_2s_3s_4s_6 - s_2s_3s_5s_6 - s_2s_3s_7 + s_2s_4 + s_2s_6 \\ s_2s_3s_4s_5 + s_2s_3s_4s_6 + s_2s_3s_5s_6 + s_2s_3s_7 \\ \vdots \\ \vdots \end{bmatrix}. \tag{5.22}$$

We observe that now we can partition $\boldsymbol{\kappa}(\boldsymbol{s})$ into $\boldsymbol{\kappa}_1(\boldsymbol{\theta}_1) = \boldsymbol{\kappa}(\boldsymbol{s})_j, j = 1, \ldots, 7$ and $\boldsymbol{\kappa}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \boldsymbol{\kappa}(\boldsymbol{s})_j, j = 8, \ldots, 31$, where

$$\boldsymbol{\theta}_1 = [s_1, s_2, s_4, s_6, s_7, s_8, s_9, s_{10}]^T$$

and

$$\boldsymbol{\theta}_2 = [s_3, s_5, s_{11}]^T.$$

As we have seen before that the derivative matrix $\boldsymbol{D}_1$ is full rank, we check $\boldsymbol{D}_2$ and find that $\boldsymbol{D}_2$ is full rank with rank 3. Therefore $\boldsymbol{D}(\boldsymbol{s})$ is full rank by Theorem 4.4.1. We conclude that the reparameterisation $\boldsymbol{s}(\boldsymbol{\theta})$ in Equation (5.21) is a simpler exhaustive summary for the model $\{\phi(t), p(t+h_C)\}$ for $K = 5, C = 2$ using Theorem 4.4.4.

For all its sub–models, we consider the determinant of the matrix U from the PLUR decomposition of the derivative matrix $\boldsymbol{D}_2$, which is given by

$$s_1^3 s_2^3 s_3^2 s_9 (s_6^2 s_{10}^2 - 2s_6^2 s_{10} + s_6 s_7 s_{10} - s_7 s_{10}^2 + s_6^2 - s_6 s_7 - s_6 s_8 + s_7^2 + 2s_7 s_{10} - s_8 s_{10} - s_7 + s_8).$$

We check all the sub–models using the following conditions (using the reparameterisation $\boldsymbol{s}(\boldsymbol{\theta})$ in Equation (5.21)):

- $\{\phi(\cdot), p(\cdot)\} : s_9 = s_1 = s_2 = s_3, s_{10} = s_4 = s_5 = 0, s_{11} = s_6 = s_7 = s_8$
- $\{\phi(\cdot), p(t)\} : s_9 = s_1 = s_2 = s_3, s_{11} = s_6 = s_7 = s_8$
- $\{\phi(\cdot), p(h_C)\} : s_9 = s_1 = s_2 = s_3, s_{10} = s_4 = s_5 = 0$
- $\{\phi(\cdot), p(t + h_C)\} : s_9 = s_1 = s_2 = s_3$
- $\{\phi(t), p(\cdot)\} : s_{10} = s_4 = s_5 = 0, s_{11} = s_6 = s_7 = s_8$
- $\{\phi(t), p(t)\} : s_{11} = s_6 = s_7 = s_8$
- $\{\phi(t), p(h_C)\} : s_{10} = s_4 = s_5 = 0$

None of the conditions makes the determinant of U equal to 0. Therefore, all sub–models of the model $\{\phi(t), p(t + h_C)\}$ in terms of the reparameterisation $\boldsymbol{s}(\boldsymbol{\theta})$ in Equation (5.21) are full rank. It follows that $\boldsymbol{s}(\boldsymbol{\theta})$ in Equation (5.21) is an exhaustive summary for $K = 5, C = 2$ for all sub–models of $\{\phi(\cdot), p(t+h_C)\}$ using Theorem 4.4.4.

Therefore we conclude that $\boldsymbol{s}(\boldsymbol{\theta})$ in Equation (5.14) is a simpler exhaustive summary for model $\{\phi(t), p(t+h_C)\}$ for $K \geq 4, C \geq 2$ using the extension theorem, Theorem 4.4.1 and the same is true of all of its sub–models for $K \geq 4, C \geq 2$ using the Theorem 4.4.2. □

5.1.3 Sub–Models of the model $\{\phi(t), p(t+h_C)\}$

Example 5.1.2. We can now use $\boldsymbol{s}(\boldsymbol{\theta})$ in Equation (5.14) as an exhaustive summary for the model $\{\phi(t), p(t+h_C)\}$ and all of its sub–models for $K \geq 4, C \geq 2$. This is simpler in both structure and length than the original exhaustive summary. We start with $K = 4, C = 2$. The new exhaustive summary is given by

$$\boldsymbol{\kappa}(\boldsymbol{\theta}) = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \tau_{p3} \\ \tau_{p4} \\ w_1\mu_p + (1-w_1)(\mu_p + \eta_{p2}) \\ w_1\mu_p^2 + (1-w_1)(\mu_p + \eta_{p2})^2 \\ w_1\mu_p^3 + (1-w_1)(\mu_p + \eta_{p2})^3 \end{bmatrix}. \tag{5.23}$$

The derivative matrix $\boldsymbol{D}(\boldsymbol{\theta})$ has rank $r_{\boldsymbol{\theta}} = 8$. Since $r_{\boldsymbol{\theta}} = n_{\boldsymbol{\theta}} = 8$, the model is full rank. To check for its sub–models, we do a PLUR decomposition of $\boldsymbol{D}(\boldsymbol{\theta})$

using Maple® , and obtain a permutation matrix

$$P = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{5.24}$$

a lower triangular matrix

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2\mu_p + 2\eta_{p2} & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 3(\mu_p + \eta_{p2})^2 & \frac{3}{2}\eta_{p2} + 3\mu_p & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{5.25}$$

a square upper triangular matrix

$$U = \begin{bmatrix} 1-w_1 & 1 & 0 & 0 & 0 & 0 & 0 & -\eta_{p2} \\ 0 & -2w_1\eta_{p2} & 0 & 0 & 0 & 0 & 0 & {\eta_{p2}}^2 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2}{\eta_{p2}}^3 \end{bmatrix}, \tag{5.26}$$

and a reduced row–echelon form matrix

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{5.27}$$

We observe that R is an identity matrix, and hence is always defined and the determinant of U is

$$\det(U) = -(1-w_1)w_1\eta_{p2}^4. \tag{5.28}$$

Recall that η_{p2} is the heterogeneous component of the recapture probability p of group 2. So the solutions that make $\det(U)$=0 are $w_1 = 0, w_1 = 1$ or $\eta_{p2} = 0$. These are only valid when there is only a single group, viz $C = 1$. This is when we have the sub–model $\{\phi(t), p(t)\}$ (the CJS model).

Hence, we conclude that the model $\{\phi(t), p(t + h_C)\}$ and all of its sub–models except the model $\{\phi(t), p(t)\}$ are full rank when $K = 4, C = 2$ using Theorem 4.4.2. This is expected as we know that for the model $\{\phi(t), p(t)\}$ (CJS), we cannot estimate the parameters ϕ_{K-1} and p_K separately, but only as a product. □

Remark 5.1.1. We can apply the same algebra to derive a simpler exhaustive summary for the model $\{\phi(t + h_C), p(t)\}$ for $K \geq 4, C \geq 2$ and for all of its sub–models.

Further, we note that the simpler exhaustive summaries for JS mixture models have the same structure as the CJS mixture models, with extra terms of $\{\beta_j, j = 0, \ldots, K - 1\}$ in the simpler exhaustive summaries.

We give examples for the models $\{\phi(t+h_C), p(t)\}$ and $\{\beta(t), \phi(t+h_C), p(t)\}$ in the relevant Maple® worksheet.

5.1.4 Model: $\{[\phi(h), p(h)]_C\}$

We now consider the case that both the capture probabilities and the survival probabilities are heterogeneous with C groups. Using the linear link, for group c at time j, we have $\phi_{jc} = \mu_\phi + \tau_{\phi j} + \eta_{\phi c}$ and $p_{jc} = \mu_p + \tau_{pj} + \eta_{pc}$, where μ describes the baseline from group 1, τ describes the time effect and η describes the heterogeneity effect.

Example 5.1.3. A possible reparameterisation for this model is given by

$$\boldsymbol{s}(\boldsymbol{\theta}) = \begin{bmatrix} \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})(\mu_p + \eta_{pc}) \\ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^2(\mu_p + \eta_{pc}) \\ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^2(\mu_p + \eta_{pc})^2 \\ \vdots \\ \vdots \\ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^{K-1}(\mu_p + \eta_{pc})^{K-1} \end{bmatrix}. \tag{5.29}$$

Consider $K = 3, C = 1$, the original exhaustive summary is given by

$$\boldsymbol{\kappa}(\boldsymbol{\theta}) = \begin{bmatrix} 1 \\ 1 - \mu_\phi + \mu_\phi\,(1 - \mu_p) \\ \mu_\phi \mu_p \\ 1 - \mu_\phi + \mu_\phi\,(1 - \mu_p)\,(1 - \mu_\phi) + {\mu_\phi}^2\,(1 - \mu_p)^2 \\ {\mu_\phi}^2\,(1 - \mu_p)\,\mu_p \\ \mu_\phi \mu_p\,(1 - \mu_\phi) + {\mu_\phi}^2\,(1 - \mu_p)\,\mu_p \\ {\mu_\phi}^2 {\mu_p}^2 \end{bmatrix}. \tag{5.30}$$

The reparameterisation is

$$\boldsymbol{s}(\boldsymbol{\theta}) = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} \mu_\phi \mu_p \\ \mu_\phi^2 \mu_p \\ \mu_\phi^2 \mu_p^2 \end{bmatrix}. \tag{5.31}$$

The reparameterisation is more complicated than it needs to be by construction in this simple case. However, this is only used as an illustration and we will extend the model to more complicated cases, for example $C = 2$.

Rewriting $\boldsymbol{\kappa}(\boldsymbol{\theta})$ using $\boldsymbol{s}(\boldsymbol{\theta})$, we obtain

$$\boldsymbol{\kappa}(\boldsymbol{s}) = \begin{bmatrix} 1 \\ 1 - s_1 \\ s_1 \\ -s_2 + s_3 + 1 - s_1 \\ -s_3 + s_2 \\ -s_3 + s_1 \\ s_3 \end{bmatrix}. \tag{5.32}$$

We find $\boldsymbol{D}(\boldsymbol{s})$ and its rank $r_{\boldsymbol{s}} = 3$. Since we have $n_{\boldsymbol{s}} = r_{\boldsymbol{s}}$, we conclude that $\boldsymbol{s}(\boldsymbol{\theta})$ in Equation (5.31) is a simpler exhaustive summary using Theorem 4.4.4.

Extending to $C = 2$ and keeping $K = 3$, the reparameterisation is

$$\boldsymbol{s}(\boldsymbol{\theta}) = \begin{bmatrix} w_1\mu_\phi\mu_p + (1 - w_1)(\mu_\phi + \eta_{\phi 2})(\mu_p + \eta_{p2}) \\ w_1\mu_\phi^2\mu_p + (1 - w_1)(\mu_\phi + \eta_{\phi 2})^2(\mu_p + \eta_{p2}) \\ w_1\mu_\phi^2\mu_p^2 + (1 - w_1)(\mu_\phi + \eta_{\phi 2})^2(\mu_p + \eta_{p2})^2 \end{bmatrix}. \tag{5.33}$$

Rewriting $\boldsymbol{\kappa}(\boldsymbol{\theta})$ using $\boldsymbol{s}(\boldsymbol{\theta})$, we note that $\boldsymbol{\kappa}(\boldsymbol{s})$ is identical to the one in Equation (5.32). It follows that $\boldsymbol{s}(\boldsymbol{\theta})$ in Equation (5.33) is a simpler exhaustive summary using Theorem 4.4.4.

Extending to $K = 4$ and keeping $C = 1$, the reparameterisation is

$$\boldsymbol{s}(\boldsymbol{\theta}) = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \end{bmatrix} = \begin{bmatrix} \mu_\phi\mu_p \\ \mu_\phi^2\mu_p \\ \mu_\phi^2\mu_p^2 \\ \mu_\phi^3\mu_p \\ \mu_\phi^3\mu_p^2 \\ \mu_\phi^3\mu_p^3 \end{bmatrix}. \tag{5.34}$$

We obtain $\boldsymbol{\kappa}(\boldsymbol{s}) = [\boldsymbol{\kappa}_1(\boldsymbol{s}_1), \boldsymbol{\kappa}_2(\boldsymbol{s}_2)]^T$, where

$$\boldsymbol{\kappa}_1(\boldsymbol{s}_1) = \begin{bmatrix} 1 \\ 1 - s_1 \\ s_1 \\ -s_2 + s_3 + 1 - s_1 \\ -s_3 + s_2 \\ -s_3 + s_1 \\ s_3 \end{bmatrix}, \tag{5.35}$$

with $\boldsymbol{s}_1 = [s_1, s_2, s_3]^T$ and

$$\boldsymbol{\kappa}_2(\boldsymbol{s}_2) = \begin{bmatrix} -s_2 + s_3 - s_4 + 2s_5 - s_6 + 1 - s_1 \\ -2s_5 + s_6 + s_4 \\ -s_3 - s_5 + s_6 + s_2 \\ -s_6 + s_5 \\ -s_3 - s_5 + s_6 + s_1 \\ -s_6 + s_5 \\ -s_6 + s_3 \\ s_6 \end{bmatrix}, \tag{5.36}$$

with $\boldsymbol{s}_2 = [s_1, s_2, s_3, s_4, s_5, s_6]^T$.

As we have seen before that $\boldsymbol{D}_1$ is full rank, we find the derivative matrix $\boldsymbol{D}_2$ has full rank 6. It follows that the model is full rank using Theorem 4.4.1. We therefore conclude that $\boldsymbol{s}(\boldsymbol{\theta})$ in Equation (5.34) is a simpler exhaustive summary using Theorem 4.4.4. Also we conclude that $\boldsymbol{s}(\boldsymbol{\theta})$ in Equation (5.29) is a simpler exhaustive summary for the model $\{[\phi(h), p(h)]_C\}$ for $K \geq 3$ and $C \geq 1$ using Theorem 4.4.1. □

5.1.5 Model: $\{[\phi(h), p(t+h)]_C\}$

We now consider the model $\{[\phi(h), p(t+h)]_C\}$. This is the previous model with a time varying component on the capture probabilities described by τ_{pj}.

Example 5.1.4. A possible reparameterisation is given by

$$\boldsymbol{s}(\boldsymbol{\theta}) = \begin{bmatrix} \tau_{p3} \\ \vdots \\ \tau_{pK} \\ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c}) \\ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})(\mu_p + \eta_{pc}) \\ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^2 \\ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^2(\mu_p + \eta_{pc}) \\ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^2(\mu_p + \eta_{pc})^2 \\ \vdots \\ \vdots \\ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^{K-1} \\ \vdots \\ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^{K-1}(\mu_p + \eta_{pc})^{K-1} \end{bmatrix}. \tag{5.37}$$

Table 5.2 summarises the estimable parameter combinations for $K = 3$ to $K = 6$ for any $C \geq 1$. We again note that $\boldsymbol{\kappa}(\boldsymbol{s})$ does not change with C as the reparameterisation $\boldsymbol{s}(\boldsymbol{\theta})$ absorbs the heterogeneous parameters. Therefore, for any $C \geq 1$, we have the same derivative matrix $\boldsymbol{D}(\boldsymbol{s})$ and hence the same rank $r_{\boldsymbol{s}}$. We note that for $K \geq 4$, there is a pattern.

By matching the $\{s_i\}$s with each of their definitions at different values of K and C, we try the new reparameterisation $\boldsymbol{s}(\boldsymbol{\theta})$ given by (5.38).

	$n_{\boldsymbol{s}}$	$r_{\boldsymbol{s}}$	Estimable parameters	
$K=3$	6	4	s_3	$s_1 s_2$ $s_1 s_4 + s_5$ $s_1 s_5 + s_6$
$K=4$	11	10	$s_1, \ldots, s_7$	$s_2 s_8 + s_9$ $-s_2^2 s_8 + s_{10}$ $s_2^3 s_8 + s_{11}$
$K=5$	17	16	$s_1, \ldots, s_{12}$	$s_3 s_{13} + s_{14}$ $-s_3^2 s_{13} + s_{15}$ $s_3^3 s_{13} + s_{16}$ $-s_3^4 s_{13} + s_{17}$
$K=6$	24	23	$s_1, \ldots, s_{18}$	$s_4 s_{19} + s_{20}$ $-s_4^2 s_{19} + s_{21}$ $s_4^3 s_{19} + s_{22}$ $-s_4^4 s_{19} + s_{23}$ $s_4^5 s_{19} + s_{24}$

Table 5.2: Summary of estimable parameters in terms of $\boldsymbol{s}(\boldsymbol{\theta})$ for $K = 3, 4, 5, 6$ for the model $\{[\phi(h), p(t+h)]_C\}$ for any $C \geq 1$.

$$\begin{bmatrix} \tau_{p3} \\ \vdots \\ \tau_{pK} \\ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c}) \\ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})(\mu_p + \eta_{pc}) \\ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^2 \\ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^2(\mu_p + \eta_{pc}) \\ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^2(\mu_p + \eta_{pc})^2 \\ \vdots \\ \vdots \\ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^{K-2}(\mu_p + \eta_{pc})^{K-2} \\ \tau_{pK} \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^{K-1} + \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^{K-1}(\mu_p + \eta_{pc}) \\ \vdots \\ (-1)^K \tau_{pK}^{K-1} \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^{K-1} + \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^{K-1}(\mu_p + \eta_{pc})^{K-1} \end{bmatrix}. \tag{5.38}$$

We further prove that this is a simpler exhaustive summary for the model $\{[\phi(h), p(t+h)]_C\}$ for $K \geq 4$ and $C \geq 1$. For full details, see the relevant Maple® supplementary materials. □

5.1.6 Model: $\{\phi(h_C), \lambda(t)\}$

We now consider a capture–recovery model $\{\phi(h_C), \lambda(t)\}$. The survival probabilities are heterogeneous and modelled using the linear link function with $\phi_{jc} = \mu_\phi + \eta_{\phi c}$, where μ_ϕ is the baseline from group 1 and $\eta_{\phi c}$ is the heterogeneous component. The recovery probabilities are time varying, denoted by $\lambda(t)$.

Example 5.1.5. A possible reparameterisation is given by

$$\boldsymbol{s}(\boldsymbol{\theta}) = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_{K-1} \\ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c}) \\ \vdots \\ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^{K-1} \end{bmatrix}. \tag{5.39}$$

Table 5.3 shows the capture histories, the probabilities of each history and the estimable parameter combinations for each $K = 2, \ldots, 5$.

$K=2$	$K=3$	$K=4$	$K=5$
		Histories	
12	012	0012	00012
	102	0102	00102
	120	0120	00120
		1002	01002
		1200	01020
			01200
			10002
			10020
			10200
			12000
		Probabilities	
$s_1 - s_1s_2$	$s_2 - s_2s_3$	$s_3 - s_3s_4$	$s_4 - s_4s_5$
	$s_2s_3 - s_2s_4$	$s_3s_4 - s_3s_5$	$s_4s_5 - s_4s_6$
	$s_1 - s_1s_3$	$s_2 - s_2s_4$	$s_3 - s_3s_5$
		$s_3s_5 - s_3s_6$	$s_4s_6 - s_4s_7$
		$s_2s_4 - s_2s_5$	$s_3s_5 - s_3s_6$
		$s_1 - s_1s_4$	$s_2 - s_2s_5$
			$s_4s_7 - s_4s_8$
			$s_3s_6 - s_3s_7$
			$s_2s_5 - s_2s_6$
			$s_1 - s_1s_5$
		Estimable parameter combinations	
	s_2/s_1	s_2/s_1	s_2/s_1
		s_3/s_1	s_3/s_4
			s_4/s_1
$(s_2-1)s_1$	$(s_4-1)s_1$	$(s_6-1)s_1$	$(s_8-1)s_1$
	$(s_3-1)s_1$	$(s_5-1)s_1$	$(s_7-1)s_1$
		$(s_4-1)s_1$	$(s_6-1)s_1$
			$(s_5-1)s_1$

Table 5.3: Summary for $K = 2, \ldots, 5$ for model $\{\phi(h_C), \lambda(t)\}$ for any $C \geq 1$.

It is sensible to try a new reparameterisation given by

$$\boldsymbol{s}(\boldsymbol{\theta}) = \begin{bmatrix} \lambda_2/\lambda_1 \\ \vdots \\ \lambda_{K-1}/\lambda_1 \\ \lambda_1 \left(-1 + \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})\right) \\ \vdots \\ \lambda_1 \left(-1 + \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^{K-1}\right) \end{bmatrix}. \tag{5.40}$$

Further we can prove that reparameterisation in Equation (5.40) is a simpler exhaustive summary for the model $\{\phi(h_C), \lambda(t)\}$ for $K \geq 2$ and $C \geq 1$, see Maple® worksheet. □

5.1.7 Model: $\{\phi(h_C + a), \lambda(t)\}$ (with covariates)

As discussed in Pledger and Schwarz (2002), we can introduce a covariate into the survival probabilities via the linear link function $\phi_{jc} = \mu_\phi + \alpha_a + \eta_{\phi c}$. We allow a potentially lower survival probability in the first year after initial capture, by setting $a = 1$ for the first year, $a = 2$ for the following years and a constraint $\alpha_1 = 0$.

This model might be appropriate for when animals are marked shortly after birth and there is higher mortality in the first year of life, compared with later years.

Example 5.1.6. A possible reparameterisation is given by

$$\boldsymbol{s}(\boldsymbol{\theta}) = \begin{bmatrix} \alpha_2 \\ \lambda_1 \\ \vdots \\ \lambda_{K-1} \\ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c}) \\ \vdots \\ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^{K-1} \end{bmatrix}. \tag{5.41}$$

Note that there is only one extra term α_2 by itself, compared to the reparameterisation in Equation (5.39) when there is no covariate considered in the model.

Table 5.4 shows the $\boldsymbol{\kappa}(\boldsymbol{s})$ for each $K = 2, \ldots, 5$. For each of these $\boldsymbol{\kappa}(\boldsymbol{s})$, we can determine the estimable parameter combinations by solving systems of partial differential equations. By considering the estimable parameter combinations in terms of $\boldsymbol{s}(\boldsymbol{\theta})$, we can derive a new reparameterisation as given below

$$\boldsymbol{u}(\boldsymbol{\theta}) = \begin{bmatrix} \lambda_2/\lambda_1 \\ \vdots \\ \lambda_{K-1}/\lambda_1 \\ \lambda_1 \left[-1 + \sum_{j=1}^{1} \binom{1}{j-1} \alpha_2^{1-j} \left\{ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^j \right\} \right] \\ \vdots \\ \lambda_1 \left[-1 + \sum_{j=1}^{K-1} \binom{K-1}{j-1} \alpha_2^{K-1-j} \left\{ \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^j \right\} \right] \end{bmatrix}. \tag{5.42}$$

Table 5.5 shows the estimable parameter combinations (left column) and the new exhaustive summary (right column) in terms of the new reparameterisation given in Equation (5.42). We can further show that the reparameterisa-

$K = 2$
$s_2 - s_2 s_3$
$K = 3$
$s_3 - s_3 s_4$
$s_3 s_4 - s_1 s_3 s_4 - s_3 s_5$
$s_2 - s_2 s_4$
$K = 4$
$s_4 - s_4 s_5$
$s_4 s_5 - s_1 s_4 s_5 - s_4 s_6$
$s_3 - s_3 s_5$
$s_1 s_4 s_5 - s_1^2 s_4 s_5 + s_4 s_6 - 2 s_1 s_4 s_6 - s_4 s_7$
$s_3 s_5 - s_1 s_3 s_5 - s_3 s_6$
$s_2 - s_2 s_5$
$K = 5$
$s_5 - s_5 s_6$
$s_5 s_6 - s_1 s_5 s_6 - s_5 s_7$
$s_4 - s_4 s_6$
$s_1 s_5 s_6 - s_1^2 s_5 s_6 + s_5 s_7 - 2 s_1 s_5 s_7 - s_5 s_8$
$s_4 s_6 - s_1 s_4 s_6 - s_4 s_7$
$s_3 - s_3 s_6$
$s_1^2 s_5 s_6 - s_1^3 s_5 s_6 + 2 s_1 s_5 s_7 - 3 s_1^2 s_5 s_7 + s_5 s_8 - 3 s_1 s_5 s_8 - s_5 s_9$
$s_1 s_4 s_6 - s_1^2 s_4 s_6 + s_4 s_7 - 2 s_1 s_4 s_7 - s_4 s_8$
$s_3 s_6 - s_1 s_3 s_6 - s_3 s_7$
$s_2 - s_2 s_6$

Table 5.4: The elements of $\boldsymbol{\kappa}(\boldsymbol{s})$ for $K = 2, \ldots, 5$, for model $\{\phi(h_C + a), \lambda(t)\}$ for any $C \geq 1$.

tion in (5.42) is a simpler exhaustive summary for this model $\{\phi(h_C+a), \lambda(t)\}$ for $K \geq 2, C \geq 2$; see the Maple® worksheet. □

Estimable parameter combinations	$\boldsymbol{\kappa}(\boldsymbol{u})$
$K = 2$	
$s_2(s_3 - 1)$	$-u_1$
$K = 3$	
s_3/s_2	$-u_1u_2$
$s_2(s_4 - 1)$	$u_1u_2 - u_1u_3$
$s_2(s_4s_1 + s_5 - 1)$	$-u_2$
$K = 4$	
$s_3/s_2, s_4/s_2$	$-u_2u_3$
$s_2(s_5 - 1)$	$u_2u_3 - u_2u_4$
$s_2(s_5s_1 + s_6 - 1)$	$-u_1u_3$
$s_2(s_5s_1^2 + 2s_1s_6 + s_7 - 1)$	$u_2u_4 - u_2u_5$
	$u_1u_3 - u_1u_4$
	$-u_3$
$K = 5$	
$s_3/s_2, s_4/s_2, s_5/s_2$	$-u_3u_4$
$s_2(s_6 - 1)$	$u_3u_4 - u_3u_5$
$s_2(s_6s_1 + s_7 - 1)$	$-u_2u_4$
$s_2(s_6s_1^2 + 2s_1s_7 + s_8 - 1)$	$u_3u_5 - u_3u_6$
$s_2(s_6s_1^3 + 3s_1s_7 + 3s_1s_8 + s_9 - 1)$	$u_2u_4 - u_2u_5$
	$-u_1u_4$
	$u_3u_6 - u_3u_7$
	$u_2u_5 - u_2u_6$
	$u_1u_4 - u_1u_5$
	$-u_4$

Table 5.5: Estimable parameter combinations and simpler exhaustive summaries for $K = 2, \ldots, 5$ for model $\{\phi(h_C + a), \lambda(t)\}$ for any $C \geq 1$.

5.1.8 Introducing the $H(x, y, z)$ function

As models get more complicated, we define

$$H(x, y, z) = \sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^x(\mu_p + \eta_{pc})^y(\mu_\lambda + \eta_{\lambda c})^z, \tag{5.43}$$

as a convenient shorthand. Recall that for a heterogeneous parameter δ, μ_δ describes the baseline from group 1 and η_δ describes the heterogeneous effect. Constraints are $\eta_{\phi 1} = \eta_{p2} = \eta_{\lambda 1} = 0$ and $w_C = 1 - \sum_{c=1}^{C-1} w_c$. We will see the use of this shorthand in later sections.

5.1.9 Model: $\{\phi(h_C), p(h_C), \lambda(h_C)\}$

In this section, we consider an example from capture–recapture and recovery. The model is constant in time and heterogeneous in all parameters.

Example 5.1.7. By trial and error, and rearranging algebraic expressions, detailed study showed that a possible reparameterisation is given by

$$H(x, y, z), \quad \text{for} \quad y = (1 - z), \ldots, x, \quad x = (1 - z), \ldots, (K - 1), \quad z = 0, 1,$$

excluding the last term $H(K - 1, K - 1, 1)$.

For each K, we have $K(K - 1)/2$ estimable $\{s_i\}$s, $K - 1$ terms of the form $X - s_1$ and $(K - 1)(K - 2)/2$ terms of the form $s_m - s_n$. Table 5.6 shows the terms which cannot be estimated separately, with Table 5.7 showing the same summary in terms of $H(x, y, z)$. From this table, we can derive a simpler exhaustive summary given by

$$\begin{cases} H(x, y, 0) & y = 1, \ldots, x, \qquad x = 1, \ldots, K - 1 \\ H(x, 0, 1) - H(0, 0, 1) & x = 1, \ldots, K - 1 \\ H(x, y, 1) - H(y, y, 1) & y = 1, \ldots, x - 1, \qquad x = 2, \ldots, K - 1. \end{cases}$$

For detailed algebraic calculations, see the Maple® worksheet.

□

$K=2$	$K=3$	$K=4$	$K=5$	$K=6$
$s_2 - s_1$	$s_2 - s_1$	$s_2 - s_1$	$s_2 - s_1$	$s_2 - s_1$
	$s_4 - s_1$	$s_4 - s_1$	$s_4 - s_1$	$s_4 - s_1$
		$s_7 - s_1$	$s_7 - s_1$	$s_7 - s_1$
			$s_{11} - s_1$	$s_{11} - s_1$
				$s_{16} - s_1$
	$s_8 - s_7$	$s_{15} - s_{13}$	$s_{24} - s_{21}$	$s_{35} - s_{31}$
		$s_{12} - s_{11}$	$s_{20} - s_{18}$	$s_{30} - s_{27}$
		$s_{14} - s_{11}$	$s_{23} - s_{18}$	$s_{34} - s_{27}$
			$s_{17} - s_{16}$	$s_{26} - s_{24}$
			$s_{19} - s_{16}$	$s_{29} - s_{24}$
			$s_{22} - s_{16}$	$s_{33} - s_{24}$
				$s_{23} - s_{22}$
				$s_{25} - s_{22}$
				$s_{28} - s_{22}$
				$s_{32} - s_{22}$

Table 5.6: Terms that cannot be estimated separately from the model $\{\phi(h_C), p(h_C), \lambda(h_C)\}$ for $K = 2, \ldots, 6$

$K=2$	$K=3$	$K=4$	$K=5$	$K=6$
$H(1,0,1)-H(0,0,1)$	$H(1,0,1)-H(0,0,1)$	$H(1,0,1)-H(0,0,1)$	$H(1,0,1)-H(0,0,1)$	$H(1,0,1)-H(0,0,1)$
	$H(2,0,1)-H(0,0,1)$	$H(2,0,1)-H(0,0,1)$	$H(2,0,1)-H(0,0,1)$	$H(2,0,1)-H(0,0,1)$
		$H(3,0,1)-H(0,0,1)$	$H(3,0,1)-H(0,0,1)$	$H(3,0,1)-H(0,0,1)$
			$H(4,0,1)-H(0,0,1)$	$H(4,0,1)-H(0,0,1)$
				$H(5,0,1)-H(0,0,1)$
	$H(2,1,1)-H(1,1,1)$	$H(2,1,1)-H(1,1,1)$	$H(2,1,1)-H(1,1,1)$	$H(2,1,1)-H(1,1,1)$
		$H(3,1,1)-H(1,1,1)$	$H(3,1,1)-H(1,1,1)$	$H(3,1,1)-H(1,1,1)$
		$H(3,2,1)-H(2,2,1)$	$H(3,2,1)-H(2,2,1)$	$H(3,2,1)-H(2,2,1)$
			$H(4,1,1)-H(1,1,1)$	$H(4,1,1)-H(1,1,1)$
			$H(4,2,1)-H(2,2,1)$	$H(4,2,1)-H(2,2,1)$
			$H(4,3,1)-H(3,3,1)$	$H(4,3,1)-H(3,3,1)$
				$H(5,1,1)-H(1,1,1)$
				$H(5,2,1)-H(2,2,1)$
				$H(5,3,1)-H(3,3,1)$
				$H(5,4,1)-H(4,4,1)$

Table 5.7: Terms that cannot be estimated separately from the model $\{\phi(h_C), p(h_C), \lambda(h_C)\}$ for $K = 2, \ldots, 6$ (written in terms of $H(x, y, z)$

5.2 Logistic Link

5.2.1 One–to–one transformations between the linear link function and the logistic link function

If there is a one–to–one transformation between the linear link function and the logistic link function, it follows from Theorem 4.4.4 (the reparameterisation theorem) that they have the same model rank, and therefore the same parameter redundancy result. We find that if only one of the parameters in the model is either heterogeneous or time–varying, then there exists a one–to–one transformation between the two link functions. We will illustrate this through Example 5.2.1.

Example 5.2.1 (One–to–one transformation)**.** Consider a model parameter Z that is only time varying. Using the linear link function, we have

$$Z_j = \mu + \tau_j,$$

with appropriate constraints such as $\mu = 0$ or $\tau_1 = 0$. For convenience, we set $\mu = 0$. Using the logistic function, we have

$$\log\left(\frac{Z_j}{1 - Z_j}\right) = \mu' + \tau'_j \quad \Longrightarrow \quad Z_j = \frac{1}{1 + \exp(-\mu' - \tau'_j)},$$

with appropriate constraints such as $\mu' = 0$ or $\tau'_1 = 0$. For convenience, we set $\mu' = 0$.

If there is a one–to–one relationship between the two link functions, we require that

$$\frac{1}{1 + \exp(-\tau'_j)} = \tau_j,$$

for all j. The system always has a unique solution. It follows that this is a one–to–one transformation between τ_j and τ'_j.

We now consider a model parameter Z that is only heterogeneous. Using the linear link function, we have

$$Z_c = \mu + \eta_c,$$

with appropriate constraints such as $\mu = 0$ or $\eta_1 = 0$. For convenience, we set $\mu = 0$. Using the logistic link function, we have

$$\log\left(\frac{Z_c}{1 - Z_c}\right) = \mu' + \eta'_c \quad \Longrightarrow \quad Z_c = \frac{1}{1 + \exp(-\mu' - \eta'_c)},$$

with appropriate constraints such as $\mu' = 0$ or $\eta'_1 = 0$. For convenience, we set $\mu' = 0$. For there to be a one–to–one relationship between the two link functions, we require that

$$\frac{1}{1 + \exp(-\eta'_c)} = \eta_c,$$

for all c. The system always has a unique solution. It follows that this is a one–to–one transformation between η_c and η'_c. □

The terms in the reparameterisation and the simpler exhaustive summaries for the models using the logistic link are more complex than those using the linear link function. We will start by showing an example of a reparameterisation for a model which uses the logistic link function below.

Example 5.2.2 (Reparameterisation using the logistic link function)**.** We start by considering a reparameterisation for the model $\{\phi(\cdot), p(t + h_C)\}$. The capture probability ϕ is constant and the survival probabilities $\{p_{jc}\}$ are time varying and heterogeneous. Using the logistic link function, the $\{p_{jc}\}$ are given by

$$\log\left(\frac{p_{jc}}{1 - p_{jc}}\right) = \mu_p + \tau_{pj} + \eta_{pc}, \tag{5.44}$$

where μ_p is the baseline from group 1, $\{\tau_{pj}\}$ are the time components and $\{\eta_{pc}\}$ are the heterogeneous components.

By trial and error, we find a possible reparameterisation is given by carrying out the following steps:

Step 1 Define a set $S_p = \{p_{jc}, j = 2, \ldots, K\}$.

Step 2 We find the set $S_p^* = 2^{S_p} \setminus \varnothing$, where 2^{S_p} is the power set of S_p and $\varnothing$ is the empty set. The number of elements in the set S_p^* is given by $\left|S_p^*\right| = 2^{|S_p|} - 1 = 2^{K-1} - 1$.

Step 3 Now we define the set

$$S_p^{**} = \left\{ x \middle| x = w_c \prod_{y \in z} y, \forall z \in S_p^* \right\}.$$

The set is formed by taking the products of each of the elements in the set S_p^*, and multiplying the products by w_c.

Step 4 The reparameterisation is given by terms in the set S, where

$$S = \{\phi\} \cup \left\{ x \middle| x = \sum_{c=1}^{C} y, y \in S_p^{**} \right\}.$$

Note that S_p, S_p^*, S_p^{**} all depends on c for $c = 1, \ldots, C$.

For $K = 3, C = 2$ for an example:

$$S_p = \{p_{2c}, p_{3c}\}$$

$$S_p^* = \{\{p_{2c}\}, \{p_{3c}\}, \{p_{2c}, p_{3c}\}\}$$

$$S_p^{**} = \{w_c p_{2c}, w_c p_{3c}, w_c p_{2c} p_{3c}\}$$

$$S = \left\{ \phi, \sum_{c=1}^{2} w_c p_{2c}, \sum_{c=1}^{2} w_c p_{3c}, \sum_{c=1}^{2} w_c p_{2c} p_{3c} \right\}$$

The final reparameterisation is given by

$$\left\{\phi,\ \sum_{c=1}^{2}\frac{w_c}{1+\exp\{-(\mu_p+\tau_{p2}+\eta_{pc})\}},\ \sum_{c=1}^{2}\frac{w_c}{1+\exp\{-(\mu_p+\tau_{p3}+\eta_{pc})\}},\ \sum_{c=1}^{2}\frac{w_c}{[1+\exp\{-(\mu_p+\tau_{p2}+\eta_{pc})\}]\,[1+\exp\{-(\mu_p+\tau_{p3}+\eta_{pc})\}]}\right\}.$$

Note that we have constraints $\tau_{p2}=0, \eta_{p1}=0$ and $w_C = 1-\sum_{c=1}^{C-1} w_c$. Therefore we can rewrite the reparameterisation in a more conventional form as a vector

$$\boldsymbol{s}(\boldsymbol{\theta}) = \begin{bmatrix} \phi \\ \frac{w_1}{1+\mathrm{e}^{-\mu_p}} + \frac{1-w_1}{1+\mathrm{e}^{-(\mu_p+\eta_{p2})}} \\ \frac{w_1}{1+\mathrm{e}^{-(\mu_p+\tau_{p3})}} + \frac{1-w_1}{1+\mathrm{e}^{-(\mu_p+\tau_{p3}+\eta_{p2})}} \\ \frac{w_1}{(1+\mathrm{e}^{-\mu_p})(1+\mathrm{e}^{-(\mu_p+\tau_{p3})})} + \frac{1-w_1}{\left\{1+\mathrm{e}^{-(\mu_p+\eta_{p2})}\right\}\left\{1+\mathrm{e}^{-(\mu_p+\tau_{p3}+\eta_{p2})}\right\}} \end{bmatrix}.$$

Recall that for the capture probabilities, μ_p describes the baseline from group 1, τ_{pj} describes the time effect and η_{pc} describes the heterogeneous effect. □

5.2.2 Model: $\{\phi(\cdot), p(t+h_C)\}$

Example 5.2.3. We now consider finding a simpler exhaustive summary for the model $\{\phi(\cdot), p(t+h_C)\}$. The survival probabilities are constant and the recapture probabilities are time varying and heterogeneous. We start with $K=3, C=2$, the original exhaustive summary is given by

$$\boldsymbol{\kappa}(\boldsymbol{\theta}) = \begin{bmatrix} 1 \\ \sum_{c=1}^{2} w_c \left\{ 1 - \phi + \phi \left(1 - \frac{1}{1+\mathrm{e}^{-(\mu_p+\tau_{p3}+\eta_{pc})}} \right) \right\} \\ \sum_{c=1}^{2} w_c \frac{\phi}{1+\mathrm{e}^{-(\mu_p+\tau_{p3}+\eta_{pc})}} \\ \vdots \\ \sum_{c=1}^{2} w_c \left[\frac{\phi(1-\phi)}{1+\mathrm{e}^{-(\mu_p+\eta_{pc})}} + \frac{\phi^2}{1+\mathrm{e}^{-(\mu_p+\eta_{pc})}} \left\{ 1 - \frac{1}{1+\mathrm{e}^{-(\mu_p+\tau_{p3}+\eta_{pc})}} \right\} \right] \\ \sum_{c=1}^{2} w_c \frac{\phi^2}{\left\{1+\mathrm{e}^{-(\mu_p+\eta_{pc})}\right\}\left\{1+\mathrm{e}^{-(\mu_p+\tau_{p3}+\eta_{pc})}\right\}} \end{bmatrix}. \quad (5.45)$$

Recall that for the recapture probabilities, μ_p is the baseline from group 1, τ_{pj} are the time components and η_{pc} are the heterogeneous components. We reparameterise using

$$\boldsymbol{s}(\boldsymbol{\theta}) = \begin{bmatrix} \phi \\ \frac{w_1}{1+\mathrm{e}^{-\mu_p}} + \frac{1-w_1}{1+\mathrm{e}^{-(\mu_p+\eta_{p2})}} \\ \frac{w_1}{1+\mathrm{e}^{-(\mu_p+\tau_{p3})}} + \frac{1-w_1}{1+\mathrm{e}^{-(\mu_p+\tau_{p3}+\eta_{p2})}} \\ \frac{w_1}{(1+\mathrm{e}^{-\mu_p})(1+\mathrm{e}^{-(\mu_p+\tau_{p3})})} + \frac{1-w_1}{\left\{1+\mathrm{e}^{-(\mu_p+\eta_{p2})}\right\}\left\{1+\mathrm{e}^{-(\mu_p+\tau_{p3}+\eta_{p2})}\right\}} \end{bmatrix}, \quad (5.46)$$

to obtain the new exhaustive summary

$$\boldsymbol{\kappa}(\boldsymbol{s}) = \begin{bmatrix} 1 \\ -s_1 s_3 + 1 \\ s_1 s_3 \\ 1 + (s_4 - s_3)s_1^2 - s_2 s_1 \\ -(s_4 - s_3)s_1^2 \\ -s_1^2 s_4 + s_1 s_2 \\ s_4 s_1^2 \end{bmatrix}. \quad (5.47)$$

We find that $\boldsymbol{D}(\boldsymbol{s})$ is full rank with $r_{\boldsymbol{s}} = 4$. We note that the dimension

of $\boldsymbol{s}(\boldsymbol{\theta})$ is $n_{\boldsymbol{s}} = 4$. It follows that $\boldsymbol{s}(\boldsymbol{\theta})$ is a simpler exhaustive summary using Theorem 4.4.4.

For $K = 3$ and any $C \geq 2$, we have the same $\boldsymbol{\kappa}(\boldsymbol{s})$. Therefore the derivative matrix $\boldsymbol{D}(\boldsymbol{s})$ is always full rank. We now consider $K = 4, C = 2$. The new reparameterisation is $\boldsymbol{s}(\boldsymbol{\theta})$ is given by

$$\begin{bmatrix} \phi \\ \frac{w_1}{1+\mathrm{e}^{-\mu_p}} + \frac{1-w_1}{1+\mathrm{e}^{-(\mu_p+\eta_{p2})}} \\ \frac{w_1}{1+\mathrm{e}^{-(\mu_p+\tau_{p3})}} + \frac{1-w_1}{1+\mathrm{e}^{-(\mu_p+\tau_{p3}+\eta_{p2})}} \\ \frac{w_1}{1+\mathrm{e}^{-(\mu_p+\tau_{p4})}} + \frac{1-w_1}{1+\mathrm{e}^{-(\mu_p+\tau_{p4}+\eta_{p2})}} \\ \frac{w_1}{(1+\mathrm{e}^{-\mu_p})(1+\mathrm{e}^{-(\mu_p+\tau_{p3})})} + \frac{1-w_1}{\left\{1+\mathrm{e}^{-(\mu_p+\eta_{p2})}\right\}\left\{1+\mathrm{e}^{-(\mu_p+\tau_{p3}+\eta_{p2})}\right\}} \\ \frac{w_1}{(1+\mathrm{e}^{-\mu_p})(1+\mathrm{e}^{-(\mu_p+\tau_{p4})})} + \frac{1-w_1}{\left\{1+\mathrm{e}^{-(\mu_p+\eta_{p2})}\right\}\left\{1+\mathrm{e}^{-(\mu_p+\tau_{p4}+\eta_{p2})}\right\}} \\ \frac{w_1}{(1+\mathrm{e}^{-(\mu_p+\tau_{p3})})(1+\mathrm{e}^{-(\mu_p+\tau_{p4})})} + \frac{1-w_1}{\left\{1+\mathrm{e}^{-(\mu_p+\tau_{p3}+\eta_{p2})}\right\}\left\{1+\mathrm{e}^{-(\mu_p+\tau_{p4}+\eta_{p2})}\right\}} \\ \frac{w_1}{(1+\mathrm{e}^{-\mu_p})(1+\mathrm{e}^{-(\mu_p+\tau_{p3})})(1+\mathrm{e}^{-(\mu_p+\tau_{p4})})} + \frac{1-w_1}{\left\{1+\mathrm{e}^{-\mu_p}\right\}\left\{1+\mathrm{e}^{-(\mu_p+\tau_{p3}+\eta_{p2})}\right\}\left\{1+\mathrm{e}^{-(\mu_p+\tau_{p4}+\eta_{p2})}\right\}} \end{bmatrix}. \tag{5.48}$$

By rewriting $\boldsymbol{\kappa}(\boldsymbol{\theta})$ using $\boldsymbol{s}(\boldsymbol{\theta})$, we obtain

$$\boldsymbol{\kappa}(\boldsymbol{s}) = \begin{bmatrix} 1 \\ -s_1 s_4 + 1 \\ s_1 s_4 \\ 1 + (s_7 - s_4)s_1^2 - s_3 s_1 \\ -(s_7 - s_4)s_1^2 \\ -s_1^2 s_7 + s_1 s_3 \\ s_7 s_1^2 \\ 1 + (-s_4 + s_6 + s_7 - s_8)s_1^3 + (s_5 - s_3)s_1^2 - s_2 s_1 \\ (s_4 + s_8 - s_6 - s_7)s_1^3 \\ (s_8 - s_7)s_1^3 + (s_3 - s_5)s_1^2 \\ (s_7 - s_8)s_1^3 \\ (s_8 - s_6)s_1^3 - s_5 s_1^2 + s_2 s_1 \\ (s_6 - s_8)s_1^3 \\ -s_1^3 s_8 + s_1^2 s_5 \\ s_8 s_1^3 \end{bmatrix}. \tag{5.49}$$

We find that the derivative matrix $\boldsymbol{D}(\boldsymbol{s})$ is full rank with $r_{\boldsymbol{s}} = 8$. We note that the dimension of $\boldsymbol{s}(\boldsymbol{\theta})$ is $n_{bs} = 8$. It follows that $\boldsymbol{s}(\boldsymbol{\theta})$ is a simpler exhaustive summary using Theorem 4.4.4.

Hence by the extension Theorem 4.4.1, the reparameterisation in Example 5.2.2 is a simpler exhaustive summary for the model $\{\phi(\cdot), p(t + h_C)\}$ for $K \geq 3, C \geq 2$. □

5.2.3 Model: $\{\phi(t), p(t + h_C)\}$

We now consider a model where both the capture probabilities and the survival probabilities are time varying with the survival probabilities also being heterogeneous.

Reparameterisation

The reparameterisation $\boldsymbol{s}(\boldsymbol{\theta})$ is given by the set S, where

$$\begin{aligned} S_p &= \{p_{jc}, j = 2, \ldots, K\}, \\ S_p^* &= 2^{S_p} \setminus \varnothing, \\ S_p^{**} &= \left\{ x \middle| x = w_c \textstyle\prod_{y \in z} y, \forall z \in S_p^* \right\}, \\ S &= \{\phi_j, j = 1, \ldots, K-1\} \cup \left\{ x \middle| x = \textstyle\sum_{c=1}^{C} y, y \in S_p^{**} \right\}. \end{aligned} \tag{5.50}$$

Note that if we have constant capture probability ϕ, this is then identical to the reparameterisation for the model $\{\phi(\cdot), p(t + h_C)\}$ as given in the previous section. As before, S_p, S_p^*, S_p^{**} all depends on c for $c = 1, \ldots, C$.

Simpler exhaustive summary

For $K = 3, C = 2$, the reparameterisation is

$$\boldsymbol{s}(\boldsymbol{\theta}) = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \frac{w_1}{1+\mathrm{e}^{-\mu_p}} + \frac{1-w_1}{1+\mathrm{e}^{-(\mu_p+\eta_{p2})}} \\ \frac{w_1}{1+\mathrm{e}^{-(\mu_p+\tau_{p3})}} + \frac{1-w_1}{1+\mathrm{e}^{-(\mu_p+\tau_{p3}+\eta_{p2})}} \\ \frac{w_1}{(1+\mathrm{e}^{-\mu_p})(1+\mathrm{e}^{-(\mu_p+\tau_{p3})})} + \frac{1-w_1}{\left\{1+\mathrm{e}^{-(\mu_p+\eta_{p2})}\right\}\left\{1+\mathrm{e}^{-(\mu_p+\tau_{p3}+\eta_{p2})}\right\}} \end{bmatrix}.$$

We obtain

$$\boldsymbol{\kappa}(\boldsymbol{s}) = \begin{bmatrix} 1 \\ -s_2 s_4 + 1 \\ s_2 s_4 \\ 1 - s_3 s_1 + (s_5 - s_4) s_2 s_1 \\ -(s_5 - s_4) s_2 s_1 \\ -s_1 s_2 s_5 + s_1 s_3 \\ s_1 s_2 s_5 \end{bmatrix}.$$

We find the derivative matrix is not full rank with $r_{\boldsymbol{s}} = 4$. The estimable parameter combinations are

$$\left\{s_1, s_3, s_2 s_4, s_2 s_5\right\}.$$

We note that this is the case for any $C \geq 2$ given $K = 3$. We now consider $K = 4, C = 2$. The reparameterisation $\boldsymbol{s}(\boldsymbol{\theta})$ is given by

$$\begin{bmatrix}
\phi_1 \\
\phi_2 \\
\phi_3 \\
\frac{w_1}{1+\mathrm{e}^{-\mu_p}} + \frac{1-w_1}{1+\mathrm{e}^{-(\mu_p+\eta_{p2})}} \\
\frac{w_1}{1+\mathrm{e}^{-(\mu_p+\tau_{p3})}} + \frac{1-w_1}{1+\mathrm{e}^{-(\mu_p+\tau_{p3}+\eta_{p2})}} \\
\frac{w_1}{1+\mathrm{e}^{-(\mu_p+\tau_{p4})}} + \frac{1-w_1}{1+\mathrm{e}^{-(\mu_p+\tau_{p4}+\eta_{p2})}} \\
\frac{w_1}{(1+\mathrm{e}^{-\mu_p})(1+\mathrm{e}^{-(\mu_p+\tau_{p3})})} + \frac{1-w_1}{\left\{1+\mathrm{e}^{-(\mu_p+\eta_{p2})}\right\}\left\{1+\mathrm{e}^{-(\mu_p+\tau_{p3}+\eta_{p2})}\right\}} \\
\frac{w_1}{(1+\mathrm{e}^{-\mu_p})(1+\mathrm{e}^{-(\mu_p+\tau_{p4})})} + \frac{1-w_1}{\left\{1+\mathrm{e}^{-(\mu_p+\eta_{p2})}\right\}\left\{1+\mathrm{e}^{-(\mu_p+\tau_{p4}+\eta_{p2})}\right\}} \\
\frac{w_1}{(1+\mathrm{e}^{-(\mu_p+\tau_{p3})})(1+\mathrm{e}^{-(\mu_p+\tau_{p4})})} + \frac{1-w_1}{\left\{1+\mathrm{e}^{-(\mu_p+\tau_{p3}+\eta_{p2})}\right\}\left\{1+\mathrm{e}^{-(\mu_p+\tau_{p4}+\eta_{p2})}\right\}} \\
\frac{w_1}{(1+\mathrm{e}^{-\mu_p})(1+\mathrm{e}^{-(\mu_p+\tau_{p3})})(1+\mathrm{e}^{-(\mu_p+\tau_{p4})})} + \frac{1-w_1}{\left\{1+\mathrm{e}^{-\mu_p}\right\}\left\{1+\mathrm{e}^{-(\mu_p+\tau_{p3}+\eta_{p2})}\right\}\left\{1+\mathrm{e}^{-(\mu_p+\tau_{p4}+\eta_{p2})}\right\}}
\end{bmatrix}.$$

We now obtain

$$\boldsymbol{\kappa}(\boldsymbol{s}) = \begin{bmatrix} 1 \\ -s_3 s_6 + 1 \\ s_3 s_6 \\ 1 - s_2 s_5 + (s_9 - s_6) s_3 s_2 \\ (s_6 - s_9) s_3 s_2 \\ -s_2 s_3 s_9 + s_2 s_5 \\ s_2 s_3 s_9 \\ 1 - s_4 s_1 + (-s_6 + s_8 + s_9 - s_{10}) s_3 s_2 s_1 + (s_7 - s_5) s_2 s_1 \\ (s_6 + s_{10} - s_8 - s_9) s_3 s_2 s_1 \\ (s_{10} - s_9) s_3 s_2 s_1 + (s_5 - s_7) s_2 s_1 \\ (s_9 - s_{10}) s_3 s_2 s_1 \\ s_4 s_1 + (s_{10} - s_8) s_3 s_2 s_1 - s_7 s_2 s_1 \\ (s_8 - s_{10}) s_3 s_2 s_1 \\ -s_1 s_2 s_3 s_{10} + s_1 s_2 s_7 \\ s_1 s_2 s_3 s_{10} \end{bmatrix}.$$

We find that the derivative matrix is not full rank with $r_{\boldsymbol{s}} = 9$. The estimable parameter combinations are given as

$$\left\{ s_1, s_2, s_4, s_5, s_7, s_3 s_6, s_3 s_8, s_3 s_9, s_3 s_{10} \right\}.$$

We can further find that the derivative matrix for $K = 5, C = 2$ has rank 18 and the estimable parameter combinations are as given below,

$$\left\{ s_1, s_2, s_3, s_5, s_6, s_7, s_9, s_{10}, s_{12}, s_{15}, s_4 s_8, s_4 s_{11}, s_4 s_{13}, s_4 s_{14}, s_4 s_{16}, s_4 s_{17}, s_4 s_{18}, s_4 s_{19} \right\}.$$

5.3 Relationship to Parameter Redundancy

This section summarises all the simpler exhaustive summaries we have found on both the linear link and the logistic link. We then discuss their relationship to the results of parameter redundancy.

5.3.1 Tables of Simpler Exhaustive Summaries

We present the simpler exhaustive summaries based on models using the linear link function in Table 5.8. Tabulated are eight Cormack–Jolly–Seber (CJS) mixture models of which only one of the two sets of parameters are heterogeneous. The model is either heterogeneous in the survival probabilities ϕ_{jc} or heterogeneous in the recapture probabilities p_{jc}. The notation $\{g(j)\}_m^n$ is used to describe the set

$$\{g(m), g(m+1), \ldots, g(n)\}$$

as a shorthand.

We observe that apart from the set of parameters that is heterogeneous, the other set of non–heterogeneous parameters can be separated from the simpler exhaustive summaries. It follows that the non–heterogeneous parameters can always be estimated separately provided there is no missing data.

These simpler exhaustive summaries can be extended to the Jolly–Seber (JS) mixture models. Since the parameters $\{\beta_j\}$ are not heterogeneous and they appear in the model by multiplying identical terms from the CJS model, $\{\beta_j\}$ can be factorised out.

For example, the simpler exhaustive summary for the CJS mixture model $\{\phi(\cdot), p(h_C)\}$ is

$$\phi, \left\{\sum_{c=1}^{C} w_c(\mu_p + \eta_{pc})^j\right\}_1^{K-1},$$

Model	Simpler Exhaustive Summary	Length	Exists for	Extend to JS?
$\{\phi(\cdot), p(h_C)\}$	$\phi, \left\{\sum_{c=1}^{C} w_c(\mu_p + \eta_{pc})^j\right\}_1^{K-1}$	K	$C \geq 1, K \geq 3$	Yes
$\{\phi(\cdot), p(t + h_C)\}$	$\phi, \{\tau_{pj}\}_3^K, \left\{\sum_{c=1}^{C} w_c(\mu_p + \eta_{pc})^j\right\}_1^{K-1}$	$2K-2$	$C \geq 1, K \geq 3$	Yes
$\{\phi(t), p(h_C)\}$	$\{\phi_j\}_1^{K-1}, \left\{\sum_{c=1}^{C} w_c(\mu_p + \eta_{pc})^j\right\}_1^{K-1}$	$2K-2$	$C \geq 1, K \geq 3$	Yes
$\{\phi(t), p(t + h_C)\}$	$\{\phi_j\}_1^{K-1}, \{\tau_{pj}\}_3^K, \left\{\sum_{c=1}^{C} w_c(\mu_p + \eta_{pc})^j\right\}_1^{K-1}$	$3K-4$	$C \geq 2, K \geq 4$	Yes
$\{\phi(h_C), p(\cdot)\}$	$p, \left\{\sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^j\right\}_1^{K-1}$	K	$C \geq 1, K \geq 3$	Yes
$\{\phi(h_C), p(t)\}$	$\{p_j\}_2^K, \left\{\sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^j\right\}_1^{K-1}$	$2K-2$	$C \geq 1, K \geq 3$	Yes
$\{\phi(t + h_C), p(\cdot)\}$	$p, \{\tau_{\phi j}\}_2^{K-1}, \left\{\sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^j\right\}_1^{K-1}$	$2K-2$	$C \geq 1, K \geq 3$	Yes
$\{\phi(t + h_C), p(t)\}$	$\{p_j\}_2^K, \{\tau_{\phi j}\}_2^{K-1}, \left\{\sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^j\right\}_1^{K-1}$	$3K-4$	$C \geq 2, K \geq 4$	Yes

Table 5.8: Simpler exhaustive summaries (CJS mixture, linear link): heterogeneity exists only in ϕ or p.

for $C \geq 1, K \geq 3$. The simpler exhaustive summary for the corresponding JS mixture model $\{\beta(t), \phi(\cdot), p(h_C)\}$ is given by

$$\{\beta_j\}_0^{K-2}, \phi, \left\{\sum_{c=1}^{C} w_c(\mu_p + \eta_{pc})^j\right\}_1^{K-1},$$

for $C \geq 1, K \geq 3$. In fact, we can show that for $\{\beta(t), \phi(\cdot), p(h_C)\}$, the simpler exhaustive summary is valid for $C \geq 1, K \geq 2$.

5.3.2 Tables of Parameter Redundancy Results

We will use an example to see how we can use simpler exhaustive summaries to determine parameter redundancy.

Example 5.3.1. Recall the simpler exhaustive summary for the CJS mixture model $\{\phi(\cdot), p(h_C)\}$ is

$$\phi, \left\{\sum_{c=1}^{C} w_c(\mu_p + \eta_{pc})^j\right\}_1^{K-1}.$$

The model is constant in the survival probability ϕ and heterogeneous in the recapture probability p_C with C mixtures. There is a single parameter ϕ for survival, C parameters for recapture and $C-1$ parameters for the weights w_c.

There are $2C$ parameters in total. The simpler exhaustive summary has length K. Since there are only K pieces of information from a K–year study and we need to estimate $2C$ parameters, it follows that a necessary condition for the model to be full rank is $K \geq 2C$. □

Example 5.3.2. Let us now consider another example. The JS mixture model $\{\beta(t), \phi(\cdot), p(t + h_C)\}$ has $2K + 2C - 3$ parameters. The simpler exhaustive summary is given by

$$\{\beta_j\}_0^{K-2}, \phi, \{\tau_{pj}\}_3^K, \left\{\sum_{c=1}^{C} w_c(\mu_p + \eta_{pc})^j\right\}_1^{K-1},$$

which has $3K-3$ terms. As there are $3K-3$ equations (from the simpler exhaustive summary) and $2K+2C-3$ parameters, a necessary condition for the model to be full rank is therefore $3K-3 \geq 2K+2C-3$, or equivalently $K \geq 2C$. □

Note. In fact, as long as heterogeneity only occur in one of the survival, recapture, or recovery probabilities, we only have to consider the parameters that are heterogeneous. For example, if a model is only heterogeneous in the survival probabilities (such as models $\{\phi(h_C), p(\cdot)\}, \{\phi(h_C), p(t)\}, \{\phi(t+h_C), p(\cdot)\}, \{\phi(t+h_C), p(t)\}, \{\phi(h_C), p(\cdot), \lambda(t)\}$ ect.), the term of interest is

$$\left\{\sum_{c=1}^{C} w_c(\mu_\phi + \eta_{\phi c})^j\right\}_1^{K-1}.$$

It has length $K-1$ and $2C-1$ model parameters. That is, we have $K-1$ equations with $2C-1$ unknowns. It follows that a necessary condition for the system to be solvable is $K-1 \geq 2C-1$, or equivalently $K \geq 2C$. For models that are only heterogeneous in the recapture probabilities, the same result follows.

Table 5.9 and Table 5.10 show just this result. When $K \geq 2C$, models are full rank and are not parameter redundant. When $K < 2C$, models are not full rank and are parameter redundant with deficiency $2C-K$.

We note that when the models are parameter redundant, $\boldsymbol{s}(\boldsymbol{\theta})$ are in fact reduced–form exhaustive summaries. That is, when $K \geq 2C$, we have $\boldsymbol{D}(\boldsymbol{s}) = \partial \boldsymbol{s}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ being full rank. It follows that $\boldsymbol{s}(\boldsymbol{\theta})$ are estimable parameters when the models are parameter redundant.

The simpler exhaustive summaries for the models when the logistic link function is used are slightly more complicated than those models using the linear link function. For example in Section 5.1.2 we have shown that for the model $\{\phi(\cdot), p(t+h_C)\}$, when the linear link function is used, the simpler

Model	Condition	Number of Estimable parameters if condition is			
		Not satisfied	(deficiency)	Satisfied	(deficiency)
$\{\phi(\cdot), p(h_C)\}$	$K \geq \max\{2C, 3\}, C \geq 1$	K	$2C - K$	$2C$	0
$\{\phi(\cdot), p(t + h_C)\}$	$K \geq \max\{2C, 3\}, C \geq 1$	$2K - 2$	$2C - K$	$K + 2C - 2$	0
$\{\phi(t), p(h_C)\}$	$K \geq \max\{2C, 3\}, C \geq 1$	$2K - 2$	$2C - K$	$K + 2C - 2$	0
$\{\phi(t), p(t + h_C)\}$	$K = 3, C \geq 2$			4	$2C - 2$
$\{\phi(t), p(t + h_C)\}$	$K \geq 2C, C \geq 2$	$3K - 4$	$2C - K$	$2K + 2C - 4$	0
$\{\phi(h_C), p(\cdot)\}$	$K \geq \max\{2C, 3\}, C \geq 1$	K	$2C - K$	$2C$	0
$\{\phi(h_C), p(t)\}$	$K \geq \max\{2C, 3\}, C \geq 1$	$2K - 2$	$2C - K$	$K + 2C - 2$	0
$\{\phi(t + h_C), p(\cdot)\}$	$K \geq \max\{2C, 3\}, C \geq 1$	$2K - 2$	$2C - K$	$K + 2C - 2$	0
$\{\phi(t + h_C), p(t)\}$	$K = 3, C \geq 2$			4	$2C - 2$
$\{\phi(t + h_C), p(t)\}$	$K \geq 2C, C \geq 2$	$3K - 4$	$2C - K$	$2K + 2C - 4$	0

Table 5.9: Parameter redundancy results (CJS mixture models with linear link function): heterogeneity exists only in ϕ or p

Model	Condition	Number of Estimable parameters if condition is			
		Not satisfied	(deficiency)	Satisfied	(deficiency)
$\{\beta(t), \phi(\cdot), p(h_C)\}$	$K \geq 2C, C \geq 1$	$2K-1$	$2C-K$	$K+2C-1$	0
$\{\beta(t), \phi(\cdot), p(t+h_C)\}$	$K \geq 2C, C \geq 1$	$3K-3$	$2C-K$	$2K+2C-3$	0
$\{\beta(t), \phi(t), p(h_C)\}$	$K \geq 2C, C \geq 1$	$3K-3$	$2C-K$	$2K+2C-3$	0
$\{\beta(t), \phi(t), p(t+h_C)\}$	$K = 2, C \geq 1$			3	$2C-2$
$\{\beta(t), \phi(t), p(t+h_C)\}$	$K = 3, C \geq 1$			6	$2C-2$
$\{\beta(t), \phi(t), p(t+h_C)\}$	$K \geq 2C, C \geq 2$	$4K-5$	$2C-K$	$3K+2C-5$	0
$\{\beta(t), \phi(h_C), p(\cdot)\}$	$K \geq 2C, C \geq 1$	$2K-1$	$2C-K$	$K+2C-1$	0
$\{\beta(t), \phi(h_C), p(t)\}$	$K \geq 2C, C \geq 1$	$3K-3$	$2C-K$	$2K+2C-3$	0
$\{\beta(t), \phi(t+h_C), p(\cdot)\}$	$K \geq 2C, C \geq 1$	$3K-3$	$2C-K$	$2K+2C-3$	0
$\{\beta(t), \phi(t+h_C), p(t)\}$	$K \geq 2C, C \geq 1$	$4K-5$	$2C-K$	$3K+2C-5$	0

Table 5.10: Parameter redundancy results (JS mixture models with linear link function): heterogeneity exists only in ϕ or p

exhaustive summary is given by the terms in the second row of Table 5.8. However, when the logistic link function is used, we found that the simpler exhaustive summary is given by the steps listed in Example 5.2.2 and proof to follow in Section 5.2.2. The simpler exhaustive summary is still simpler than the original exhaustive summary, but it is more complicated to describe mathematically or to write using a set of expressions. The same complications apply to the model $\{\phi(t), p(t + h_C)\}$ when the logistic link function is used in Section 5.2.3. Because of this reason, we do not provide table of exhaustive summaries for models when the logistic link function is used. This is part of the future work to come.

5.4 Conclusion

Using the symbolic method we discussed in Chapter 4, we have seen that it is possible to find the estimable parameters for parameter redundant models for small values of K and C by solving the systems of partial differential equations (PDEs). The solutions we find can be used to suggest a possible reparameterisation $\boldsymbol{s}(\boldsymbol{\theta})$, with appropriate algebraic rearrangement.

If the reparameterisation $\boldsymbol{s}(\boldsymbol{\theta})$ is a simpler exhaustive summary, we can then use $\boldsymbol{s}(\boldsymbol{\theta})$ to determine the results of parameter redundancy, by comparing the number of terms in $\boldsymbol{s}(\boldsymbol{\theta})$ and the number of model parameters. If the reparameterisation $\boldsymbol{s}(\boldsymbol{\theta})$ is not a simpler exhaustive summary, we can find a simpler exhaustive summary by solving the appropriate PDEs.

Models using the linear link function are simpler in structure than models using the logistic link function. We find that the simpler exhaustive summaries from using the linear link function are simpler than those from using the logistic link function.

Models with interaction terms between the time effect and the heterogeneous effect are far more complex than any of their sub–models. Deriving simple exhaustive summaries for such models is challenging and is part of the work to be done, in the future, as described in Chapter 7.

5.5 List of files

In this section, we list all relevant files on the supplementary CD in Table 5.11. Details on using the Maple® package `MixtureModelforCaptureRecapture.mla` are given in Appendix A and details on using the Mathematica® package `MixtureModelforCaptureRecapture.m` are given in Appendix B.

File Name	Description
`MixtureModelforCaptureRecapture.mla`	The Maple® package
`Chapter_5.1_linear.mw`	Worksheet for Section 5.1.
`Chapter_5.2_logistic.mw`	Worksheet for Section 5.2.

Table 5.11: List of files on Chapter 5

Chapter 6

Fitting Mixture Models to Open Population Data

In Chapter 4, using the SN method, we have checked the parameter redundancy result for both CJS and JS mixture models for $3 \leq K \leq 10$ and $2 \leq C \leq 10$. We note that the rule of $K \geq 2C$ can ensure all parameters are estimable. But this is too restrictive for some models, for example, the model $\{[\phi(h), p(h)]_C\}$. In Chapter 5, by developing simpler exhaustive summaries, we formally checked that the rule of $K \geq 2C$ is an exact rule for some models, but it is not an exact rule for all models to ensure that all parameters are estimable.

However, the rule of $K \geq 2C$ can guarantee that all model parameters are estimable. Without knowing the rule, we risk fitting too many mixture components and not being able to estimate the model parameters, or fearing having too many components and not fitting a plausible model. In this chapter, we will fit mixture models to real open population data with all plausible models, compare our results with the published results and demonstrate the difficulty of fitting open population mixture models.

Section 6.1 uses real datasets from Pledger *et al.* (2003). We find different

fitted results and select some models to use as examples to demonstrate the difficulty of fitting open population mixture models. We show that our fitted results are better than the results from Pledger *et al.* (2003) by comparing the maximized log–likelihood values, checking the gradient at the maximum–likelihood estimates as well as using profile likelihood (Pawitan, 2001; Royston, 2007) plots.

Section 6.2 shows we have fitted all possible models, excluding those models with interactions between time and heterogeneous components. We also state the method of model selection and show that we have found a structurally simpler model than the best models given in Pledger *et al.* (2003), by allowing up to four mixture components, while Pledger *et al.* (2003) have only fitted models with two mixture components.

The use of different link functions is discussed again in Section 6.3. We focus on the effect of model fitting rather than the parameter redundancy result as discussed in Chapter 4. We note again that different parameter redundancy results can be obtained from using the linear link function and the logistic link function if there is not a *one–to–one transformation*.

Models with interactions between the time and heterogeneous components are discussed in Section 6.4. We note that by having different constraints on the interaction matrices, the point estimates are different. Hence, a different likelihood can be obtained. It is then obvious that if we do not know what constraints have been used when we consider a model with interactions, it will be difficult to reproduce the same result. We give two examples of the fitted result of the same model by putting different constraints on the interaction matrix as a comparison.

Lastly we conclude the chapter in Section 6.5 and list all computer code used in Section 6.6.

6.1 Fitting Real Datasets

We find that open population capture–recapture mixture models are difficult to fit in general. Even with two mixture components, optimization could stop at a wrong place, see Section 6.1.3 for example. In this chapter, we will use examples to illustrate the difficulty of fitting mixture models. We refer to the paper by Pledger *et al.* (2003) using PPN2003. The authors of this paper fitted 11 models to the possum dataset, which is described below. We will compare our fitted results with the results from PPN2003.

6.1.1 Brushtail Possum (*Trichosurus vulpecula*) Data

Capture–recapture data on the Australian brushtail possum (*Trichosurus vulpecula*) have been provided by the first author in PPN2003, together with details of their model fitting results. This is a nine–year study of data collected from the Orongorongo Valley near Wellington, New Zealand, involving marking 175 female individuals and 273 male individuals. There are 57 distinct female capture histories and 58 distinct male capture histories. Overall there is a total of 448 sampled possums with 86 unique capture histories. All 9 samples are taken in February from 1980 to 1988.

6.1.2 Optimization Routines

Pledger *et al.* (2003) used the computer package `R` for all the model fitting. We have fitted the same data initially using Maple® . Since Maple® does not have a built–in global optimization routine, we have used an add–on package called `DirectSearch`, which uses a direct search algorithm (Moiseev,

2011). The Maple® code runs very slowly so we explored a different computer package Mathematica® . The reason for switching to Mathematica® is that it is capable of doing symbolic algebra as well as numerical optimization. Mathematica® offers four numerical global optimization routines, namely Nelder–Mead (Nelder and Mead, 1965), differential evolution (Storn and Price, 1995), simulated annealing (Kirkpatrick *et al.*, 1983) and random search (Zhigljavsky, 1991).

We note that the symbolic packages Maple® and Mathematica® have been used extensively for Chapters 3,4 and 5. We will keep using symbolic package in this chapter. The main reason is that before we fit the model, we still need to check that the model is not parameter redundant due to *imperfect* data. Recall that in Section 1.2 we discussed the two types of parameter redundancy, intrinsic parameter redundancy and extrinsic parameter redundancy (Hubbard, 2014). Our main focus is to determine the former, where parameter redundancy only occurs due to the model structure. The general condition $K \geq 2C$ can ensure that the model is not parameter redundant due to its model structure but it is only a guideline when fitting real datasets. It is important that we check whether the model is parameter redundant *before* we fit it. For this reason, we still need to compute the symbolic derivative matrix and calculate the symbolic rank or the numeric rank using the SN method where appropriate.

We focus on the method of random search as it searches the parameter space more thoroughly by using more starting points from the Mathematica® function `RandomReal`. For example, for each model, we have used 4000 starting points (see `CJSfit_Linear_Link.pdf`) when the linear link function is used and 500 starting points

(see `CJSfit_Logistic_Link.pdf`) when the logistic link function is used. We have allowed 500 iterations for all optimization. For some simple models, like $\{\phi(\cdot), p(h_C)\}$, 100 starting point may suffices. For more complicated models,

like $\{[\phi(t+h), p(t+h)]_C\}$, we need 200 start points or more to ensure that we have a global maximum. The main reason for that many starting points is to ensure for all models that we can find a global maximum. When the linear link function is used, some starting points do not satisfy the constraint which ensures that all point estimates are between 0 and 1. Therefore, we need far more starting points for the linear link function than the logistic link function. However, there is no guarantee. We check by comparing the maximized likelihood values, so that the bigger model always performs as well as all its sub–models. By default, Mathematica® stops the optimization if the difference between two successive likelihood values is less than 10^{-8} on 64–bit operating systems. Full code and output has been attached to the supplementary CD. For description of files, see Section 6.6.

We note that PPN2003 uses the optimization method `L-BFGS-B` (Byrd *et al.*, 1995) available in `R`, which limits the memory usage as well as the boundaries of the parameter space. This could stop the optimization from getting a global maximum, see the example in the next section for instance. PPN2003 used 5 iterations with each of a 100 cyclic fixing. The stopping criterion is that if two successive values of maximum–likelihood values are within 0.000001 of each other, the iteration stops. There is therefore no check on the gradient vectors. For more details, see supplementary file `hcjsRun`, lines 65–94.

Next, for illustration we give one example each from the fitted results on the combined data, female data and the male data. These fitted results are then compared with the results from PPN2003.

6.1.3 Example: $\{\phi(t), p(t + h_2)\}$

In model $\{\phi(t), p(t + h_2)\}$, the survival probabilities are time–varying and the recapture probabilities are time–varying and heterogeneous with two groups. That is, the recapture probabilities are modelled by

$$\log\left(\frac{p_{jc}}{1 - p_{pj}}\right) = \mu_p + \tau_{pj} + \eta_{pc},$$

for $j = 2, 3, \ldots, 9$ and $c = 1, 2$.

The following is an extract of the fitted results to the combined brushtail possum data provided by the first author in PPN2003.

possfeb.out

```
1  $phit.pth2me:
2  $phit.pth2me$info:
3      Max.ll       RD   rawAIC Npar K    Pi-hat    Pi-hat
4   -848.3808 1696.762 1732.762   18 9 0.3039709 0.6960291
5
6  $phit.pth2me$details:
7       Sample     p-hat     p-hat    mean.p   phi-hat   ann.phi
8  [1,]      1        NA        NA        NA 0.6096223 0.6096223
9  [2,]      2 0.9420076 0.9979831 0.9809681 0.7068549 0.7068549
10 [3,]      3 0.6985363 0.9860302 0.8986404 0.8264058 0.8264058
11 [4,]      4 0.3964154 0.9523944 0.7833930 0.7946019 0.7946019
12 [5,]      5 0.4683808 0.9640774 0.8134001 0.7630867 0.7630867
13 [6,]      6 0.3644067 0.9458417 0.7691023 0.8256249 0.8256249
14 [7,]      7 0.2378723 0.9048291 0.7020936 0.7864057 0.7864057
15 [8,]      8 0.2726637 0.9194800 0.7228667 0.8171094 0.8171094
16 [9,]      9 0.1667687 0.8590895 0.6486441        NA        NA
```

For this particular model, the authors constrained all model parameters on the logistic scale to be between -20 and 20.

Line 4 of the above extract gives the maximized log–likelihood value of -848.3808 with an AIC value of 1732.762. It also gives the estimate $\hat{w}_1 = 0.3039709$ or equivalently $\hat{w}_1 = 0.6960291$. Lines 6–16 give the point estimates of the two groups of capture probabilities and the survival probabilities on the natural scale.

As the recapture probabilities $\{p_{jc}\}$ are heterogeneous and given, we can use this information to work out which link function PPN2003 has used. Using

the point estimates, we find that

$$\begin{cases} \hat{p}_{22} - \hat{p}_{21} = 0.9979831 - 0.9420076 = 0.055976 \\ \hat{p}_{32} - \hat{p}_{31} = 0.9860302 - 0.6985363 = 0.287494 \end{cases},$$

and

$$\begin{cases} \log\left(\frac{\hat{p}_{22}}{1-\hat{p}_{22}}\right) - \log\left(\frac{\hat{p}_{21}}{1-\hat{p}_{21}}\right) = 3.4164 \\ \log\left(\frac{\hat{p}_{32}}{1-\hat{p}_{32}}\right) - \log\left(\frac{\hat{p}_{31}}{1-\hat{p}_{31}}\right) = 3.4164 \end{cases}.$$

Since using the logistic link function, the differences between the first and second group are the same, we deduce that the logistic link function has been used. We further deduce that the heterogeneous component for the recapture probability is given by $\hat{\eta}_p = -3.4164$.

To find the maximum–likelihood estimates of the survival probabilities $\phi_j, j = 1, \ldots, 8$, on the logistic scale, we solve the system of equations,

$$\begin{bmatrix} \frac{1}{1+\exp(-\phi_1)} \\ \frac{1}{1+\exp(-\phi_2)} \\ \frac{1}{1+\exp(-\phi_3)} \\ \frac{1}{1+\exp(-\phi_4)} \\ \frac{1}{1+\exp(-\phi_5)} \\ \frac{1}{1+\exp(-\phi_6)} \\ \frac{1}{1+\exp(-\phi_7)} \\ \frac{1}{1+\exp(-\phi_8)} \end{bmatrix} = \begin{bmatrix} 0.6096223 \\ 0.7068549 \\ 0.8264058 \\ 0.7946019 \\ 0.7630867 \\ 0.8256249 \\ 0.7864057 \\ 0.8171094 \end{bmatrix}.$$

Similarly, for the recapture probabilities we solve

$$\begin{bmatrix} \frac{1}{1+\exp(-\mu_p)} \\ \frac{1}{1+\exp(-\mu_p-\tau_3)} \\ \frac{1}{1+\exp(-\mu_p-\tau_4)} \\ \frac{1}{1+\exp(-\mu_p-\tau_5)} \\ \frac{1}{1+\exp(-\mu_p-\tau_6)} \\ \frac{1}{1+\exp(-\mu_p-\tau_7)} \\ \frac{1}{1+\exp(-\mu_p-\tau_8)} \\ \frac{1}{1+\exp(-\mu_p-\tau_9)} \end{bmatrix} = \begin{bmatrix} 0.9420076 \\ 0.6985363 \\ 0.3964154 \\ 0.4683808 \\ 0.3644067 \\ 0.2378723 \\ 0.2726637 \\ 0.1667687 \end{bmatrix}.$$

There are 18 model parameters. We use Mathematica® with the optimization method of `RandomSearch` with 200 starting points to find the maximum–likelihood estimates. A summary of our maximum–likelihood estimates and the estimates derived from the point estimates above are given in Table 6.1. We obtain two different sets of the maximum–likelihood estimates giving the same maximized log–likelihood value since we can relabel the groups. In terms of optimization, we can easily achieve this by constraining $0 \leq w \leq 0.5$ or $0.5 \leq w \leq 1$ on the natural scale, or equivalently constraining $w \leq 0$ or $w \geq 0$ on the logistic scale. Our two identical maximum log–likelihood values give -842.34, compared to -848.38 from PPN2003.

As a check that our estimates are indeed the global optimum, we produce a profile likelihood plot (Stafford and Andrews, 1993) for each of the 18 model parameters in Figure 6.1 and Figure 6.2. All model parameters except w are on the logistic scale. The solid dots indicate the values we have calculated to draw the curve.

The horizontal dashed lines indicate the position of the maximized log–likelihood values. For graphs with two horizontal dashed lines, the higher one

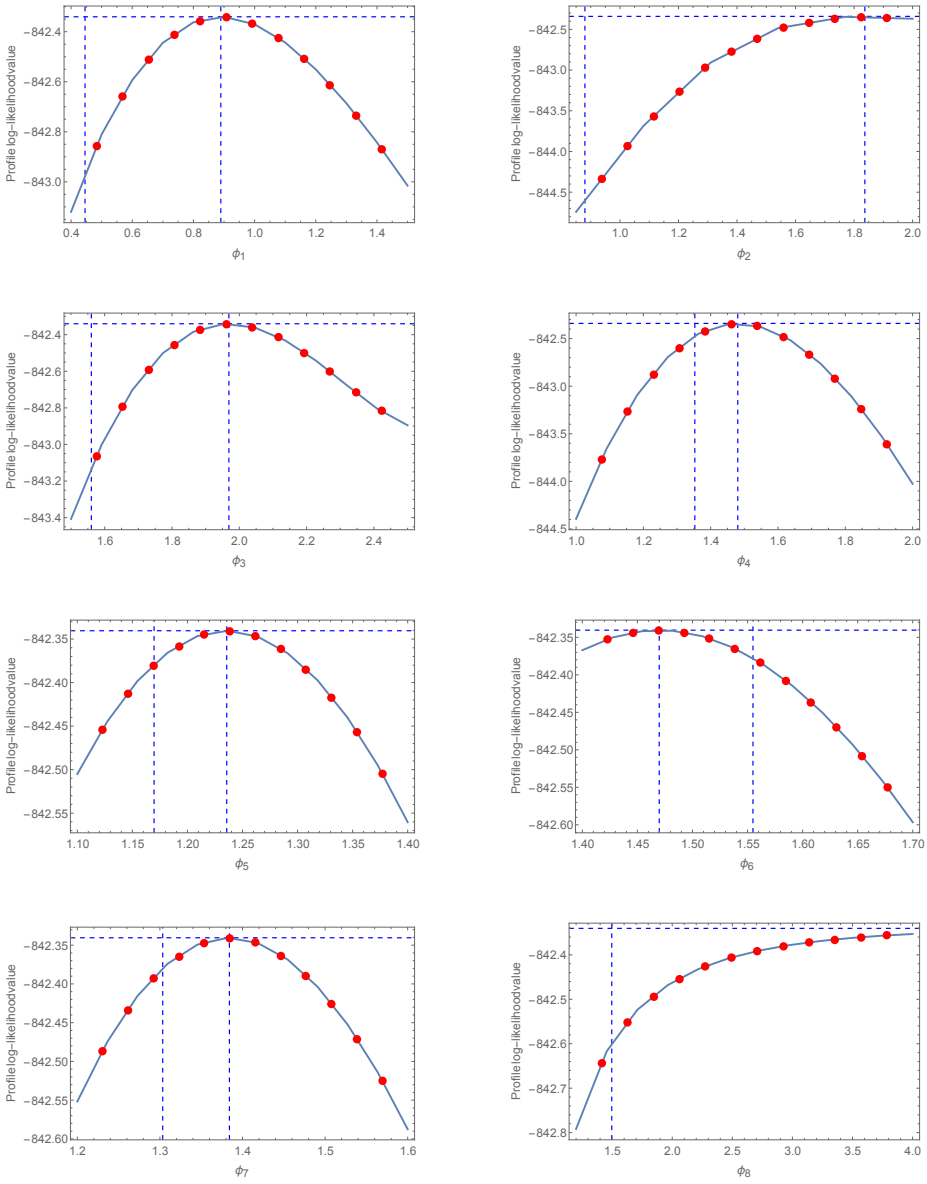

Figure 6.1: Profile plots of $\phi_j, j = 1, \ldots, 8$ for model $\{\phi(t), p(t + h_2)\}$. The vertical dotted lines indicate the positions of the maximum–likelihood estimates. The horizontal dashed lines indicate the position of the maximized log–likelihood values. Our vertical lines always meet the horizontal line on the profile curve at the point which gives the maximum log–likelihood value.

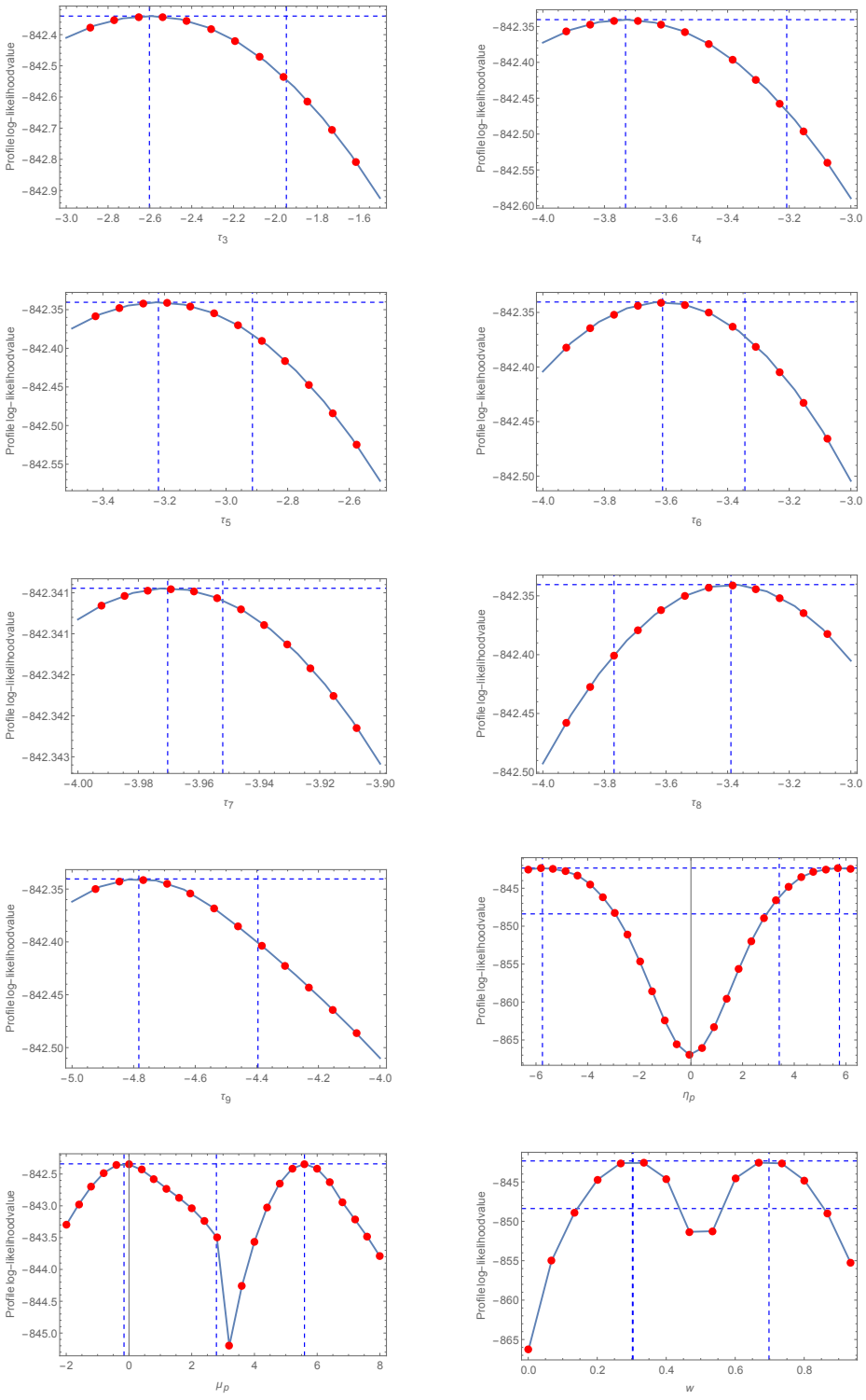

Figure 6.2: Profile plots of μ_p, η_p, w and $\tau_j, j = 3, \ldots, 9$ for model $\{\phi(t), p(t + h_2)\}$. The vertical dotted lines indicate the positions of the maximum–likelihood estimates. The horizontal dashed lines indicate the position of the maximized log–likelihood values. For graphs with two horizontal dashed lines, the higher one represents our maximized log–likelihood value and the lower one represents the maximized log–likelihood from PPN2003.

	Our result			PPN2003	
Parameter	MLE1	MLE2	Gradient	MLE	Gradient
w	0.302	0.698	0.000	0.304	2.805
η_p	5.752	−5.752	0.000	3.417	−1.882
μ_p	−0.158	5.594	0.000	2.788	−4.283
τ_3	−2.602	−2.602	0.000	−1.947	−0.233
τ_4	−3.732	−3.732	0.000	−3.208	−0.745
τ_5	−3.220	−3.220	0.000	−2.914	−0.352
τ_6	−3.611	−3.611	0.000	−3.344	−0.554
τ_7	−3.970	−3.970	0.000	−3.952	−0.902
τ_8	−3.389	−3.389	0.000	−3.769	−0.629
τ_9	−4.783	−4.783	0.000	−4.396	−0.721
ϕ_1	0.890	0.890	0.000	0.446	0.043
ϕ_2	1.837	1.837	0.000	0.880	0.541
ϕ_3	1.969	1.969	0.000	1.560	0.573
ϕ_4	1.481	1.481	0.000	1.353	0.791
ϕ_5	1.236	1.236	0.000	1.170	0.209
ϕ_6	1.470	1.470	0.000	1.555	0.086
ϕ_7	1.384	1.384	0.000	1.303	0.417
ϕ_8	18.802	14.718	0.000	1.497	0.596

Table 6.1: Estimated parameters for model $\{\phi(t), p(t + h_2)\}$. All estimates except w are on the logistic scale. All values are rounded to 4 decimal places.

indicates the position of our maximized log–likelihood value and the lower one indicates the position of the maximized log–likelihood value from PPN2003. For example, see the profile plots of η_p and w. For graphs with only one horizontal dashed line, the line always indicates our maximized log–likelihood value.

The vertical dashed lines indicate the positions of the maximum–likelihood estimates. We note that there are clear differences between our estimates and the PPN2003 estimates. From the profile plots, our estimates give a vertical line at the maximum point of each profile curve while the estimates from PPN2003 do not.

We observe that ϕ_8 is estimated on a boundary. So our estimates of $\hat{\phi}_8 = 18.8023$ and $\hat{\phi}_8 = 14.7183$ are suggesting the same information. That is, the survival probability ϕ_8 on the natural scale is very close to 1. The profile

plots for μ_p, η_p and w have two maxima as we would expect.

Table 6.2 shows the point estimates of this model. There is a clear difference between our average recapture rates and the results from PPN2003. Our p_js are lower at each j. We have a much bigger difference between the groups in the recapture probabilities than in the results from PPN2003.

	Our	PPN2003
ϕ_1	0.709	0.610
ϕ_2	0.863	0.707
ϕ_3	0.878	0.826
ϕ_4	0.815	0.795
ϕ_5	0.775	0.763
ϕ_6	0.813	0.826
ϕ_7	0.800	0.786
ϕ_8	1.000	0.817

(a) Probability of survival

Our	Group 1	Group 2	Average
p_2	0.461	0.996	0.835
p_3	0.060	0.952	0.683
p_4	0.020	0.866	0.610
p_5	0.033	0.915	0.649
p_6	0.023	0.879	0.620
p_7	0.016	0.835	0.588
p_8	0.028	0.901	0.637
p_9	0.007	0.692	0.485
PPN2003	Group 1	Group 2	Average
p_2	0.942	0.998	0.981
p_3	0.699	0.986	0.899
p_4	0.396	0.952	0.783
p_5	0.468	0.964	0.813
p_6	0.364	0.946	0.769
p_7	0.238	0.905	0.702
p_8	0.273	0.919	0.723
p_9	0.167	0.859	0.649

(b) Probability of recapture

Table 6.2: Point estimates for model $\{\phi(t), p(t + h_2)\}$. All probabilities are on natural scale and between 0 and 1.

6.1.4 Example: $\{[\phi(t+h), p(t+h)]_2\}$ (Similar & Opposite)

In this model, both the survival $\{\phi_{jc}\}$ and recapture $\{p_{jc}\}$ probabilities are time–varying and heterogeneous with two groups. The square brackets indicate that $\{\phi_{jc}\}$ and $\{p_{jc}\}$ share the same $\{w_c\}$. For both the survival and the recapture probabilities, we consider a group with higher values and denote these with a subscript 'H' and another group with lower values and denote these with a subscript 'L'. This leads to the concept of a "similar" model and an "opposite" model, see Figure 6.3.

	Similar	Opposite
w_1	$\{\phi_{\mathrm{H}}, p_{\mathrm{H}}\}$ or $\{\phi_{\mathrm{L}}, p_{\mathrm{L}}\}$	$\{\phi_{\mathrm{H}}, p_{\mathrm{L}}\}$ or $\{\phi_{\mathrm{L}}, p_{\mathrm{H}}\}$
$1 - w_1$	$\{\phi_{\mathrm{L}}, p_{\mathrm{L}}\}$ or $\{\phi_{\mathrm{H}}, p_{\mathrm{H}}\}$	$\{\phi_{\mathrm{L}}, p_{\mathrm{H}}\}$ or $\{\phi_{\mathrm{H}}, p_{\mathrm{L}}\}$

Figure 6.3: Illustration of the "similar" model and the "opposite" model.

Using the logistic link function, we model the survival probabilities through

$$\log\left(\frac{\phi_{jc}}{1-\phi_{jc}}\right) = \mu_\phi + \tau_{\phi j} + \eta_{\phi c}, \tag{6.1}$$

for $j = 1, 2, \ldots, 8$ and $c = 1, 2$, where μ_ϕ describes the baseline from group 1, $\tau_{\phi j}$ describes the time effect and $\eta_{\phi c}$ describes the heterogeneous effect. The recapture probabilities are given by

$$\log\left(\frac{p_{jc}}{1-p_{jc}}\right) = \mu_p + \tau_{pj} + \eta_{pc}, \tag{6.2}$$

for $j = 2, 3, \ldots, 9$ and $c = 1, 2$, where μ_p describes the baseline from group 1, τ_{pj} describes the time effect and η_{pc} describes the heterogeneous effect. Recall that $\eta_{\phi 1} = \eta_{p1} = 0$.

For the "similar" model, we require that $\eta_{p2} \cdot \eta_{\phi 2} > 0$. That is, η_{p2} and $\eta_{\phi 2}$ have the same sign. For the "opposite" model, we require that $\eta_{p2} \cdot \eta_{\phi 2} < 0$. That is η_{p2} and $\eta_{\phi 2}$ have opposite sign. For example for the "similar model", both the survival probabilities and the recapture probabilities in group 1, they

are either both lower than those in group 2, or they are both higher than those in group 2. But the probabilities cannot have $p_{j1} > p_{j2}$ in one of the two groups and $\phi_{j1} < \phi_{j2}$ in the other group, or $p_{j1} < p_{j2}$ in one of the two group and $\phi_{j1} > \phi_{j2}$ in the other group for all j.

We present the results of this model fitted to the female possum data as an example. Table 6.3 shows our maximized log–likelihood values ($\ell_{\max}$) and maximum–likelihood estimates, compared with those from PPN2003. By comparing the maximised log–likelihood values, we have obtained a larger maximum likelihood. Our results show that the "similar" model fits slightly better than the "opposite" model in terms of the likelihood.

We note that some maximum–likelihood estimates are large, but this is not a problem since they are on the logistic scale. Table 6.4 gives the point estimates for both the "similar" model and the "opposite" model.

We also note that in our "opposite" model, the gradient of the model parameter η_{ϕ_2} is 1.3856. We give the profile plot of $\eta_{\phi 2}$ in Figure 6.4. The plot shows clearly that the likelihood value at $\hat{\eta}_{\phi 2} = -0.000$ is smaller than the likelihood value at $\hat{\eta}_{\phi 2} = 0.7514$. By constraining on $\eta_{p2}\eta_{\phi 2} < 0$, the optimization has stopped at a boundary for $\eta_{\phi 2}$.

We suggest that there is no need to sub–divide the model $\{[\phi(t+h), p(t+h)]_2\}$ into a "similar" and an "opposite" model. If we do not add any constraints on η_{p2} and $\eta_{\phi 2}$, the optimization will identify the better model of the two.

	Similar				Opposite			
	PPN2003		Our		PPN2003		Our	
$\ell_{\max}$	-362.5302		-362.2840		-364.1337		-362.8710	
Parameter	Estimate	Gradient	Estimate	Gradient	Estimate	Gradient	Estimate	Gradient
w_1	0.468	-2.071	0.389	-0.006	0.240	0.068	0.342	0.000
η_{p2}	2.340	0.019	2.396	0.001	2.407	-0.027	2.458	0.000
$\eta_{\phi 2}$	0.812	0.250	0.751	0.001	-0.602	1.092	0.000	1.386
μ_p	5.317	0.099	20.156	0.001	5.109	-0.261	14.034	0.000
μ_ϕ	0.175	0.181	0.141	0.003	1.062	-0.561	0.578	0.000
τ_{p3}	-3.184	0.005	-18.222	-0.000	-3.193	-0.024	-12.001	0.000
τ_{p4}	-4.351	0.011	-19.409	0.001	-4.299	-0.030	-13.143	0.000
τ_{p5}	-3.787	0.003	-18.870	-0.001	-3.674	-0.040	-12.556	0.000
τ_{p6}	-4.711	0.015	-19.815	0.000	-4.607	-0.044	-13.507	0.000
τ_{p7}	-5.384	0.020	-20.449	0.000	-5.216	-0.045	-14.155	0.000
τ_{p8}	-5.262	-0.005	-20.324	-0.001	-5.306	-0.038	-14.222	0.000
τ_{p9}	-6.020	-0.014	-21.390	0.001	-6.139	-0.102	-15.271	0.000
$\tau_{\phi 2}$	0.568	-0.011	0.573	0.001	0.560	-0.104	0.571	0.000
$\tau_{\phi 3}$	1.557	0.008	1.570	0.000	1.509	-0.152	1.536	0.000
$\tau_{\phi 4}$	0.741	-0.034	0.752	0.001	0.833	-0.114	0.825	0.000
$\tau_{\phi 5}$	1.268	0.040	1.320	-0.002	1.295	-0.088	1.342	0.000
$\tau_{\phi 6}$	1.436	0.020	1.409	0.000	1.417	-0.106	1.494	0.000
$\tau_{\phi 7}$	0.609	0.016	0.635	0.000	0.947	-0.196	0.916	0.000
$\tau_{\phi 8}$	1.574	0.131	21.900	0.000	1.881	0.276	33.001	0.000

Table 6.3: Maximum–likelihood estimates for $\{[\phi(t+h), p(t+h)]_2\}$

similar	PPN2003		Our		opposite	PPN2003		Our	
	$c=1$	$c=2$	$c=1$	$c=2$		$c=1$	$c=2$	$c=1$	$c=2$
p_{2c}	0.995	1.000	1.000	1.000	p_{2c}	0.994	0.999	1.000	1.000
p_{3c}	0.894	0.989	0.874	0.987	p_{3c}	0.872	0.987	0.884	0.989
p_{4c}	0.724	0.965	0.678	0.959	p_{4c}	0.692	0.961	0.709	0.966
p_{5c}	0.822	0.980	0.783	0.975	p_{5c}	0.808	0.979	0.814	0.981
p_{6c}	0.647	0.950	0.584	0.939	p_{6c}	0.623	0.948	0.629	0.952
p_{7c}	0.483	0.907	0.427	0.891	p_{7c}	0.473	0.909	0.470	0.912
p_{8c}	0.514	0.916	0.458	0.903	p_{8c}	0.451	0.901	0.453	0.906
p_{9c}	0.331	0.837	0.225	0.762	p_{9c}	0.263	0.798	0.225	0.772
ϕ_{1c}	0.544	0.729	0.535	0.709	ϕ_{1c}	0.743	0.613	0.641	0.641
ϕ_{2c}	0.678	0.826	0.671	0.812	ϕ_{2c}	0.835	0.735	0.759	0.759
ϕ_{3c}	0.850	0.927	0.847	0.921	ϕ_{3c}	0.929	0.877	0.892	0.892
ϕ_{4c}	0.714	0.849	0.710	0.838	ϕ_{4c}	0.869	0.785	0.803	0.803
ϕ_{5c}	0.809	0.905	0.812	0.901	ϕ_{5c}	0.914	0.853	0.872	0.872
ϕ_{6c}	0.834	0.919	0.825	0.909	ϕ_{6c}	0.923	0.867	0.888	0.888
ϕ_{7c}	0.686	0.831	0.685	0.822	ϕ_{7c}	0.882	0.803	0.817	0.817
ϕ_{8c}	0.852	0.928	1.000	1.000	ϕ_{8c}	0.950	0.912	1.000	1.000

Table 6.4: Point estimates for $\{[\phi(t+h), p(t+h)]_2\}$. All values are rounded to 3 decimal places. Note that for the opposite model, our estimates for the survival probabilities are the same for both groups. This is expected as the maximum–likelihood estimate of $\hat{\eta}_{\phi 2} = -0.000$ to 3 decimal places.

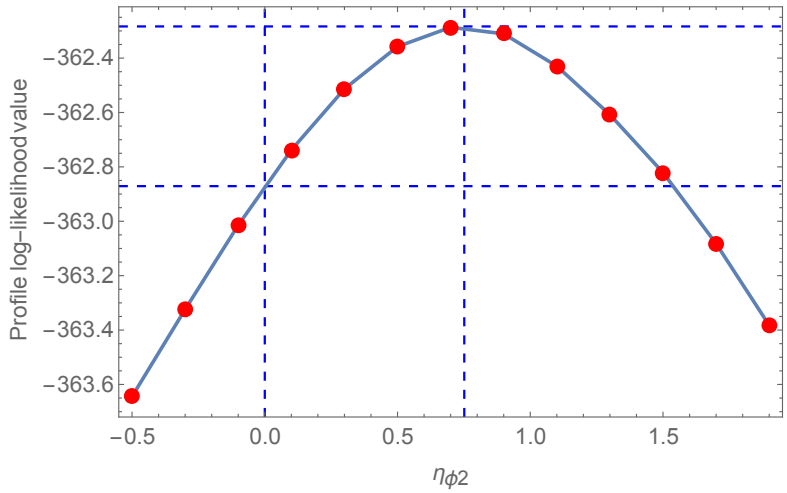

Figure 6.4: Profile plot of $\eta_{\phi 2}$ for model $\{[\phi(t+h), p(t+h)]_2\}$. The left vertical dashed line indicates the maximum–likelihood estimate from the "opposite" model and the right vertical dashed line indicates the maximum–likelihood estimate from the "similar" model. By constraining on $\eta_{\phi 2}\eta_{p2} < 0$, the optimization stopped at a boundary.

6.1.5 Example: $\{\phi(t + h_2), p(t + h_2)\}$ Four–class cross–classified model

In this model, both the survival $\{\phi_{jc}\}$ and recapture $\{p_{jc}\}$ probabilities are time–varying and heterogeneous with two groups. Unlike the previous model, $\{[\phi(t+h), p(t+h)]_2\}$, which shares the same $\{w_c\}$, the four–class cross–classified model $\{\phi(t+h_2), p(t+h_2)\}$ has separate $\{w_c\}$s for each possible group as shown in Figure 6.5. We use 'L' for the group with lower probabilities and 'H' for the group with higher probabilities. We set $w_4 = 1 - w_1 - w_2 - w_3$.

	ϕ_{jc}	p_{jc}
w_1	L	L
w_2	L	H
w_3	H	L
w_4	H	H

Figure 6.5: Illustration of the four–class cross–classified model $\{\phi(t+h_2), p(t+h_2)\}$.

We show the results of fitting this model to the male possum data as an example. Table 6.5a compares our maximum–likelihood estimates and gradients at the estimates with the results from PPN2003. We note that the results from PPN2003 did not find the global optimum. Using the same logistic link function as given in Equations (6.1) and (6.2), we obtain the point estimates in Table 6.5b.

We note that our estimated recapture probabilities p_{j2} are close to 1 and the survival probabilities ϕ_{j1} are close to 0. Both are estimated on boundaries. Together with the estimate of $\hat{w}_3 = 0.000$, this is not surprising. Recall in Figure 6.5, the combination of ϕ_{j1} and p_{j2} are the *third* group. This extreme case has a probability of being close to 0, therefore the weight that is estimated is close to 0 as well.

6.2 Model Selection

As already observed, the Brushtail Possum data is a $K = 9$ year study. Following the rule of $K \geq 2C$, we fit all possible 39 CJS mixture models, excluding those with interactions. The fitted results are given in Table 6.6 for the female possum data, Table 6.7 is for the male possum data and Table 6.8 is for the combined possum data.

In the three tables we give the maximized log–likelihood values, the AIC values, the relative AIC values (ΔAIC) and the order of the relative AIC values. We subtract the smallest AIC value from the original AIC to give ΔAIC. Then each of the 39 models are ordered according to their ΔAICs. The smallest ΔAIC has order 1 and the biggest ΔAIC has order 39. All values are rounded to 2 decimals where appropriate.

	PPN2003		Our	
$\ell_{\max}$	-461.708		-457.741	
Parameter	Estimate	Gradient	Estimate	Gradient
w_1	0.505	-0.936	0.308	0.000
w_2	0.121	2.350	0.273	0.000
w_3	0.000	-7.073	0.000	0.000
η_{p2}	8.201	0.021	19.189	0.000
$\eta_{\phi 2}$	1.568	0.105	13.166	0.000
μ_p	2.168	0.048	2.233	0.000
μ_ϕ	-0.386	-2.562	-11.472	0.000
τ_{p3}	-1.296	0.013	-1.236	0.000
τ_{p4}	-3.081	-0.051	-2.740	0.000
τ_{p5}	-2.534	0.038	-2.267	0.000
τ_{p6}	-1.903	-0.007	-1.844	0.000
τ_{p7}	-2.437	-0.002	-2.250	0.000
τ_{p8}	-1.100	0.007	-0.851	0.000
τ_{p9}	1.367	0.044	8.376	0.000
$\tau_{\phi 2}$	0.283	-0.227	-0.662	0.000
$\tau_{\phi 3}$	1.620	-0.279	0.296	0.000
$\tau_{\phi 4}$	0.552	-0.381	-0.185	0.000
$\tau_{\phi 5}$	-0.433	-0.399	-1.199	0.000
$\tau_{\phi 6}$	0.199	-0.410	-0.656	0.000
$\tau_{\phi 7}$	0.101	-0.320	-0.358	0.000
$\tau_{\phi 8}$	-0.333	-0.388	-1.004	0.000

(a) Maximum–likelihood estimates for $\{\phi(t+h_2), p(t+h_2)\}$

	PPN2003		Our	
	$c=1$	$c=2$	$c=1$	$c=2$
p_{2c}	0.897	1.000	0.903	1.000
p_{3c}	0.705	1.000	0.730	1.000
p_{4c}	0.286	0.999	0.376	1.000
p_{5c}	0.410	1.000	0.492	1.000
p_{6c}	0.566	1.000	0.596	1.000
p_{7c}	0.433	1.000	0.496	1.000
p_{8c}	0.744	1.000	0.799	1.000
p_{9c}	0.972	1.000	1.000	1.000
ϕ_{1c}	0.405	0.765	0.000	0.845
ϕ_{2c}	0.475	0.812	0.000	0.737
ϕ_{3c}	0.775	0.943	0.000	0.880
ϕ_{4c}	0.542	0.850	0.000	0.819
ϕ_{5c}	0.306	0.679	0.000	0.621
ϕ_{6c}	0.454	0.799	0.000	0.738
ϕ_{7c}	0.429	0.783	0.000	0.792
ϕ_{8c}	0.328	0.700	0.000	0.666

(b) Point estimates for $\{\phi(t+h_2), p(t+h_2)\}$

Table 6.5: Comparison of results for model $\{\phi(t+h_2), p(t+h_2)\}$

Model	$\ell_{\max}$	AIC	ΔAIC	Order
$\{\phi(\cdot), p(\cdot)\}$	-380.53	765.06	13.72	28
$\{\phi(\cdot), p(t)\}$	-374.53	767.06	15.72	35
$\{\phi(\cdot), p(h_2)\}$	-372.75	753.51	2.16	6
$\{\phi(\cdot), p(h_3)\}$	-370.48	752.96	1.62	2
$\{\phi(\cdot), p(h_4)\}$	-370.43	756.86	5.51	12
$\{\phi(\cdot), p(t+h_2)\}$	-365.58	753.15	1.81	4
$\{\phi(\cdot), p(t+h_3)\}$	-362.67	751.34	0.00	1
$\{\phi(\cdot), p(t+h_4)\}$	-362.60	755.20	3.86	8
$\{\phi(t), p(\cdot)\}$	-373.87	765.74	14.40	30
$\{\phi(t), p(h_2)\}$	-371.81	765.61	14.27	29
$\{\phi(t), p(h_3)\}$	-368.38	762.75	11.41	24
$\{\phi(t), p(h_4)\}$	-368.33	766.66	15.32	33
$\{\phi(t), p(t+h_2)\}$	-362.87	761.74	10.40	19
$\{\phi(t), p(t+h_3)\}$	-360.13	760.27	8.92	15
$\{\phi(t), p(t+h_4)\}$	-360.06	764.11	12.77	25
$\{\phi(h_2), p(\cdot)\}$	-372.75	753.51	2.16	5
$\{\phi(h_3), p(\cdot)\}$	-372.75	757.51	6.16	14
$\{\phi(h_4), p(\cdot)\}$	-372.75	761.51	10.16	18
$\{\phi(h_2), p(t)\}$	-365.58	753.15	1.81	3
$\{\phi(h_3), p(t)\}$	-365.58	757.15	5.81	13
$\{\phi(h_4), p(t)\}$	-365.58	761.15	9.81	17
$\{[\phi(h), p(h)]_2\}$	-372.75	755.51	4.16	10
$\{[\phi(h), p(h)]_3\}$	-370.28	756.56	5.22	11
$\{[\phi(h), p(h)]_4\}$	-370.22	762.44	11.10	20
$\{[\phi(h), p(t+h)]_2\}$	-365.58	755.15	3.81	7
$\{[\phi(h), p(t+h)]_3\}$	-362.67	755.34	4.00	9
$\{[\phi(h), p(t+h)]_4\}$	-362.56	761.13	9.79	16
$\{\phi(t+h_2), p(\cdot)\}$	-370.36	762.72	11.38	23
$\{\phi(t+h_3), p(\cdot)\}$	-370.36	766.72	15.38	34
$\{\phi(t+h_4), p(\cdot)\}$	-370.36	770.72	19.38	38
$\{\phi(t+h_2), p(t)\}$	-363.28	762.56	11.21	21
$\{\phi(t+h_3), p(t)\}$	-363.28	766.56	15.21	31
$\{\phi(t+h_4), p(t)\}$	-363.28	770.56	19.21	37
$\{[\phi(t+h), p(h)]_2\}$	-370.36	764.72	13.38	27
$\{[\phi(t+h), p(h)]_3\}$	-368.32	766.64	15.30	32
$\{[\phi(t+h), p(h)]_4\}$	-368.27	772.54	21.20	39
$\{[\phi(t+h), p(t+h)]_2\}$	-362.28	762.57	11.22	22
$\{[\phi(t+h), p(t+h)]_3\}$	-360.13	764.26	12.92	26
$\{[\phi(t+h), p(t+h)]_4\}$	-359.94	769.87	18.53	36

Table 6.6: Summary of model fitting of female possum data using the logistic link function. The smallest AIC is circled.

Model	$\ell_{\max}$	AIC	ΔAIC	Order
$\{\phi(\cdot), p(\cdot)\}$	-491.20	986.40	38.73	37
$\{\phi(\cdot), p(t)\}$	-487.31	992.63	44.96	38
$\{\phi(\cdot), p(h_2)\}$	-475.28	958.56	10.90	14
$\{\phi(\cdot), p(h_3)\}$	-467.83	947.67	0.00	1
$\{\phi(\cdot), p(h_4)\}$	-467.83	951.67	4.00	4
$\{\phi(\cdot), p(t+h_2)\}$	-468.76	959.52	11.86	17
$\{\phi(\cdot), p(t+h_3)\}$	-461.69	949.38	1.71	2
$\{\phi(\cdot), p(t+h_4)\}$	-461.69	953.38	5.71	7
$\{\phi(t), p(\cdot)\}$	-488.02	994.03	46.37	39
$\{\phi(t), p(h_2)\}$	-470.58	963.15	15.49	20
$\{\phi(t), p(h_3)\}$	-463.08	952.16	4.49	5
$\{\phi(t), p(h_4)\}$	-463.08	956.16	8.49	10
$\{\phi(t), p(t+h_2)\}$	-464.17	964.34	16.67	22
$\{\phi(t), p(t+h_3)\}$	-456.87	953.74	6.08	8
$\{\phi(t), p(t+h_4)\}$	-456.87	957.74	10.08	13
$\{\phi(h_2), p(\cdot)\}$	-480.13	968.26	20.60	25
$\{\phi(h_3), p(\cdot)\}$	-480.13	972.26	24.60	27
$\{\phi(h_4), p(\cdot)\}$	-480.13	976.26	28.60	30
$\{\phi(h_2), p(t)\}$	-474.73	971.46	23.80	26
$\{\phi(h_3), p(t)\}$	-474.73	975.46	27.80	29
$\{\phi(h_4), p(t)\}$	-474.73	979.46	31.80	33
$\{[\phi(h), p(h)]_2\}$	-474.67	959.35	11.68	16
$\{[\phi(h), p(h)]_3\}$	-467.73	951.47	3.80	3
$\{[\phi(h), p(h)]_4\}$	-467.73	957.47	9.80	12
$\{[\phi(h), p(t+h)]_2\}$	-468.76	961.52	13.86	18
$\{[\phi(h), p(t+h)]_3\}$	-461.65	953.31	5.64	6
$\{[\phi(h), p(t+h)]_4\}$	-461.65	959.31	11.64	15
$\{\phi(t+h_2), p(\cdot)\}$	-475.40	972.80	25.14	28
$\{\phi(t+h_3), p(\cdot)\}$	-475.40	976.80	29.14	32
$\{\phi(t+h_4), p(\cdot)\}$	-475.40	980.80	33.14	35
$\{\phi(t+h_2), p(t)\}$	-470.30	976.61	28.94	31
$\{\phi(t+h_3), p(t)\}$	-470.30	980.61	32.94	34
$\{\phi(t+h_4), p(t)\}$	-470.30	984.61	36.94	36
$\{[\phi(t+h), p(h)]_2\}$	-470.46	964.93	17.26	24
$\{[\phi(t+h), p(h)]_3\}$	-462.87	955.75	8.08	9
$\{[\phi(t+h), p(h)]_4\}$	-462.87	961.75	14.08	19
$\{[\phi(t+h), p(t+h)]_2\}$	-463.43	964.87	17.20	23
$\{[\phi(t+h), p(t+h)]_3\}$	-456.73	957.46	9.79	11
$\{[\phi(t+h), p(t+h)]_4\}$	-456.72	963.45	15.78	21

Table 6.7: Summary of model fitting of male possum data using the logistic link function. The smallest AIC is circled.

Model	$\ell_{\max}$	AIC	ΔAIC	Order
$\{\phi(\cdot), p(\cdot)\}$	-882.19	1768.39	68.60	37
$\{\phi(\cdot), p(t)\}$	-875.83	1769.67	69.88	39
$\{\phi(\cdot), p(h_2)\}$	-855.78	1719.57	19.78	17
$\{\phi(\cdot), p(h_3)\}$	-847.74	1707.48	7.69	4
$\{\phi(\cdot), p(h_4)\}$	-847.66	1711.33	11.54	8
$\{\phi(\cdot), p(t+h_2)\}$	-844.92	1711.85	12.06	9
$\{\phi(\cdot), p(t+h_3)\}$	-836.89	1699.79	0.00	1
$\{\phi(\cdot), p(t+h_4)\}$	-836.30	1702.61	2.82	2
$\{\phi(t), p(\cdot)\}$	-875.45	1768.90	69.11	38
$\{\phi(t), p(h_2)\}$	-853.61	1729.21	29.43	26
$\{\phi(t), p(h_3)\}$	-845.34	1716.68	16.89	14
$\{\phi(t), p(h_4)\}$	-845.29	1720.58	20.79	20
$\{\phi(t), p(t+h_2)\}$	-842.34	1720.68	20.89	21
$\{\phi(t), p(t+h_3)\}$	-834.88	1709.75	9.96	6
$\{\phi(t), p(t+h_4)\}$	-834.33	1712.67	12.88	10
$\{\phi(h_2), p(\cdot)\}$	-860.40	1728.79	29.00	24
$\{\phi(h_3), p(\cdot)\}$	-860.40	1732.79	33.00	28
$\{\phi(h_4), p(\cdot)\}$	-860.40	1736.79	37.00	31
$\{\phi(h_2), p(t)\}$	-851.55	1725.10	25.31	22
$\{\phi(h_3), p(t)\}$	-851.55	1729.10	29.31	25
$\{\phi(h_4), p(t)\}$	-851.55	1733.10	33.31	29
$\{[\phi(h), p(h)]_2\}$	-855.02	1720.04	20.26	19
$\{[\phi(h), p(h)]_3\}$	-847.43	1710.86	11.07	7
$\{[\phi(h), p(h)]_4\}$	-847.28	1716.57	16.78	13
$\{[\phi(h), p(t+h)]_2\}$	-844.91	1713.82	14.04	12
$\{[\phi(h), p(t+h)]_3\}$	-836.70	1703.40	3.62	3
$\{[\phi(h), p(t+h)]_4\}$	-835.99	1707.97	8.18	5
$\{\phi(t+h_2), p(\cdot)\}$	-857.98	1737.96	38.17	32
$\{\phi(t+h_3), p(\cdot)\}$	-857.98	1741.96	42.17	34
$\{\phi(t+h_4), p(\cdot)\}$	-857.98	1745.96	46.17	36
$\{\phi(t+h_2), p(t)\}$	-849.26	1734.52	34.74	30
$\{\phi(t+h_3), p(t)\}$	-849.26	1738.52	38.74	33
$\{\phi(t+h_4), p(t)\}$	-849.26	1742.52	42.74	35
$\{[\phi(t+h), p(h)]_2\}$	-852.89	1729.79	30.00	27
$\{[\phi(t+h), p(h)]_3\}$	-845.01	1720.03	20.24	18
$\{[\phi(t+h), p(h)]_4\}$	-844.86	1725.72	25.93	23
$\{[\phi(t+h), p(t+h)]_2\}$	-840.69	1719.39	19.60	16
$\{[\phi(t+h), p(t+h)]_3\}$	-834.85	1713.70	13.91	11
$\{[\phi(t+h), p(t+h)]_4\}$	-833.98	1717.97	18.18	15

Table 6.8: Summary of model fitting of combined possum data using the logistic link function. The smallest AIC is circled.

Following the AIC model selection criteria (or ΔAIC), model $\{\phi(\cdot), p(t + h_3)\}$ is selected for the female dataset, model $\{\phi(\cdot), p(h_3)\}$ is selected for the male dataset and model $\{\phi(\cdot), p(t + h_3)\}$ is selected for the combined dataset. By comparing the AIC values (see supplementary file `possfeb.pdf`) from PPN2003, our best models are better than the best models chosen in PPN2003. We have explored all plausible model by allowing up to four mixture components. This is an important finding since the models selected are conceptually simpler than the models selected in PPN2003. We also note that the models we have found all have constant survival probabilities, which also makes it easier to interpret the model.

6.3 Using Different Link Functions

Recall that in Example 5.2.1, we discussed the effect of using different link functions on the result of parameter redundancy. If there is a one–to–one transformation between the links, the same parameter redundancy results hold for both link functions. This is also true for model fitting by maximum log–likelihood (Morgan, 2008).

Table 6.9 shows the AIC values for all three sets of possum data from using both the linear link function and the logistic link function.

Models with a † indicate that there is a one–to–one transformation between the linear link function and the logistic link function. There are 18 out of the 39 models that are invariant to the choice of the link functions. We observe that if there exists a one–to–one transformation, the models will have the same AIC value as expected.

In some rare cases, for example see the values in circles, using the linear link function gives a smaller AIC value than using the logistic link function.

	Combined		Female		Male	
	logit	linear	logit	linear	logit	linear
$\{\phi(\cdot), p(\cdot)\}^\dagger$	1768	1768	765	765	986	986
$\{\phi(\cdot), p(t)\}^\dagger$	1770	1770	767	767	993	993
$\{\phi(\cdot), p(h_2)\}^\dagger$	1720	1720	754	754	959	959
$\{\phi(\cdot), p(h_3)\}^\dagger$	1707	1707	753	753	948	948
$\{\phi(\cdot), p(h_4)\}^\dagger$	1711	1711	757	757	952	952
$\{\phi(\cdot), p(t+h_2)\}$	1712	1712	753	759	960	960
$\{\phi(\cdot), p(t+h_3)\}$	1700	1712	751	761	949	960
$\{\phi(\cdot), p(t+h_4)\}$	1703	1716	755	765	953	964
$\{\phi(t), p(\cdot)\}^\dagger$	1769	1769	766	766	994	994
$\{\phi(t), p(h_2)\}^\dagger$	1729	1729	766	763	963	963
$\{\phi(t), p(h_3)\}^\dagger$	1717	1717	763	763	952	952
$\{\phi(t), p(h_4)\}^\dagger$	1721	1721	767	767	956	956
$\{\phi(t), p(t+h_2)\}$	1721	1722	762	766	964	965
$\{\phi(t), p(t+h_3)\}$	1710	1722	760	770	954	965
$\{\phi(t), p(t+h_4)\}$	1713	1726	764	773	958	969
$\{\phi(h_2), p(\cdot)\}^\dagger$	1729	1729	754	754	968	968
$\{\phi(h_3), p(\cdot)\}^\dagger$	1733	1733	758	758	972	972
$\{\phi(h_4), p(\cdot)\}^\dagger$	1737	1737	762	762	976	976
$\{\phi(h_2), p(t)\}^\dagger$	1725	1725	753	753	971	971
$\{\phi(h_3), p(t)\}^\dagger$	1729	1729	757	757	975	975
$\{\phi(h_4), p(t)\}^\dagger$	1733	1733	761	761	979	979
$\{[\phi(h), p(h)]_2\}^\dagger$	1720	1720	756	756	959	959
$\{[\phi(h), p(h)]_3\}^\dagger$	1711	1711	757	757	951	951
$\{[\phi(h), p(h)]_4\}^\dagger$	1717	1717	762	763	957	957
$\{[\phi(h), p(t+h)]_2\}$	1714	1713	755	755	962	961
$\{[\phi(h), p(t+h)]_3\}$	1703	1714	755	759	953	963
$\{[\phi(h), p(t+h)]_4\}$	1708	1720	761	765	959	969
$\{\phi(t+h_2), p(\cdot)\}$	1738	1739	763	763	973	975
$\{\phi(t+h_3), p(\cdot)\}$	1742	1743	767	767	977	979
$\{\phi(t+h_4), p(\cdot)\}$	1746	1747	771	771	981	983
$\{\phi(t+h_2), p(t)\}$	1735	1735	763	763	977	979
$\{\phi(t+h_3), p(t)\}$	1739	1739	767	767	981	983
$\{\phi(t+h_4), p(t)\}$	1743	1743	771	771	985	987
$\{[\phi(t+h), p(h)]_2\}$	1730	1730	765	765	965	965
$\{[\phi(t+h), p(h)]_3\}$	1720	1720	767	766	956	956
$\{[\phi(t+h), p(h)]_4\}$	1726	1726	773	772	962	962
$\{[\phi(t+h), p(t+h)]_2\}$	1719	1721	763	766	965	965
$\{[\phi(t+h), p(t+h)]_3\}$	1714	1723	764	771	957	967
$\{[\phi(t+h), p(t+h)]_4\}$	1718	1729	770	775	963	973

Table 6.9: Comparison of model AIC values from using logistic link and linear link. Circled values are explained in the text.

For the combined data, the model $\{[\phi(h), p(t+h)]_2\}$ using the linear link gives an AIC value of 1713 while using the logistic link give an AIC value of 1714. For the female data, models $\{[\phi(t+h), p(h)]_3\}$ and $\{[\phi(t+h), p(h)]_4\}$ show that the model with the linear link has a lower AIC value by the difference of 1 compared to the logistic link. These differences are most likely due to the rounding errors of model fitting.

Example 6.3.1 (One–to–one transformation: numerical example)**.** We now give a numerical example to show a one–to–one transformation between the linear link function and the logistic link function.

The model $\{\phi(\cdot), p(h_4)\}$ has constant survival probability and heterogeneous recapture probability with 4 groups. The maximum–likelihood estimates from using the linear link function are given below

$$\begin{cases} \hat{\phi} &=& 0.798982 \\ \hat{w}_1 &=& 0.227750 \\ \hat{w}_2 &=& 0.334073 \\ \hat{w}_3 &=& 0.397814 \\ \hat{\eta}_{p2} &=& 0.768960 \\ \hat{\eta}_{p3} &=& 0.998797 \\ \hat{\eta}_{p4} &=& 0.185628 \\ \hat{\mu}_p &=& 0.000439744 \end{cases},$$

resulting in a maximized log–likelihood value of -847.678 with point estimates of p_{jc} given by

$$\begin{cases} \hat{p}_{j1} &=& \hat{\mu}_p &=& 0.000439744 \\ \hat{p}_{j2} &=& \hat{\mu}_p + \hat{\eta}_{p2} &=& 0.769400 \\ \hat{p}_{j3} &=& \hat{\mu}_p + \hat{\eta}_{p3} &=& 0.999236 \\ \hat{p}_{j4} &=& \hat{\mu}_p + \hat{\eta}_{p4} &=& 0.186068 \end{cases},$$

for all $j = 2, 3, \ldots, 9$. We now consider using the logistic link function on the recapture probabilities to give,

$$\log\left(\frac{p_{jc}}{1-p_{jc}}\right) = \mu_p + \eta_{pc}.$$

Equivalently, we have

$$\begin{cases} p_{j1} &= \frac{1}{1+\exp(-\mu_p)} \\ p_{j2} &= \frac{1}{1+\exp(-\mu_p-\eta_{p2})} \\ p_{j3} &= \frac{1}{1+\exp(-\mu_p-\eta_{p3})} \\ p_{j4} &= \frac{1}{1+\exp(-\mu_p-\eta_{p4})} \end{cases}.$$

To get the estimates on the logistic scale, we solve

$$\begin{cases} 0.000439744 &= \frac{1}{1+\exp(-\mu_p)} \\ 0.7694 &= \frac{1}{1+\exp(-\mu_p-\eta_{p2})} \\ 0.999236 &= \frac{1}{1+\exp(-\mu_p-\eta_{p3})} \\ 0.186068 &= \frac{1}{1+\exp(-\mu_p-\eta_{p4})} \end{cases}$$

to obtain

$$\begin{cases} \hat{\eta}'_{p2} &= 8.9338 \\ \hat{\eta}'_{p3} &= 14.9056 \\ \hat{\eta}'_{p4} &= 6.25311 \\ \hat{\mu'}_p &= -7.72888 \end{cases}.$$

We obtain the estimate of ϕ on the logistic scale by solving

$$\frac{1}{1+\exp(-\phi)} = 0.798982 \quad \Longrightarrow \quad \hat{\phi}' = 1.37994.$$

The new set of estimates

$$\begin{cases} \hat{\phi}' &= 1.37994 \\ \hat{w}_1 &= 0.227750 \\ \hat{w}_2 &= 0.334073 \\ \hat{w}_3 &= 0.397814 \\ \hat{\eta}'_{p2} &= 8.9338 \\ \hat{\eta}'_{p3} &= 14.9056 \\ \hat{\eta}'_{p4} &= 6.25311 \\ \hat{\mu}'_p &= -7.72888 \end{cases},$$

are the equivalent maximum–likelihood estimates from using the logistic link function. We can double check this by substituting them into the corresponding log–likelihood and confirm that it gives the same maximum log–likelihood value of -847.678. □

When there is no one–to–one transformation between the links, in most cases the models with the logistic link function result in a lower AIC. Thus the logistic model performs better than using the linear link function and is more likely to be chosen based on AIC.

Next we give an example below where there is no one–to–one transformation between the linear link function and the logistic link function for comparison.

Example 6.3.2 (Not a one–to–one transformation: numerical example)**.** We now show an example which there is no one–to–one transformation between the linear link function and the logistic link function. Consider the model $\{\phi(\cdot), p(t + h_4)\}$: The survival probability is constant and the recapture probabilities are time–varying and there is a heterogeneous recapture probability with 4 groups. The maximum–likelihood estimates from using the linear link

function are given by

$$\begin{cases} \hat{\phi} &= 0.806564 \\ \hat{w}_1 &= 0.270547 \\ \hat{w}_2 &= 0.054746 \\ \hat{w}_3 &= 0.047751 \\ \hat{\eta}_{p2} &= 0.563662 \\ \hat{\eta}_{p3} &= 0.563662 \\ \hat{\eta}_{p4} &= 0.907902 \\ \hat{\mu}_p &= 0.092098 \\ \hat{\tau}_{p3} &= -0.04141 \\ \hat{\tau}_{p4} &= -0.0920984 \\ \hat{\tau}_{p5} &= -0.056473 \\ \hat{\tau}_{p6} &= -0.0920968 \\ \hat{\tau}_{p7} &= -0.0806524 \\ \hat{\tau}_{p8} &= -0.0840715 \\ \hat{\tau}_{p9} &= -0.0920984 \end{cases},$$

giving a maximized log–likelihood value of -843.244 with point estimates of p_{jc} given below

	$c=1$	$c=2$	$c=3$	$c=4$
$j=2$	0.0920980	0.655760	0.655760	1.000000
$j=3$	0.0506880	0.614350	0.614350	0.958590
$j=4$	0.0000000	0.563662	0.563662	0.907902
$j=5$	0.0356250	0.599287	0.599287	0.943527
$j=6$	0.0000000	0.563663	0.563663	0.907903
$j=7$	0.0114456	0.575108	0.575108	0.919348
$j=8$	0.0080265	0.571689	0.571689	0.915929
$j=9$	0.0000000	0.563662	0.563662	0.907902

.

Using the logistic link function, the recapture probabilities $\{p_{jc}\}$ are given by

$$\begin{bmatrix} \frac{1}{1+e^{-\mu_p}} & \frac{1}{1+e^{-(\mu_p+\eta_{p2})}} & \frac{1}{1+e^{-(\mu_p+\eta_{p3})}} & \frac{1}{1+e^{-(\mu_p+\eta_{p4})}} \\ \frac{1}{1+e^{-(\mu_p+\tau_{p3})}} & \frac{1}{1+e^{-(\mu_p+\tau_{p3}+\eta_{p2})}} & \frac{1}{1+e^{-(\mu_p+\tau_{p3}+\eta_{p3})}} & \frac{1}{1+e^{-(\mu_p+\tau_{p3}+\eta_{p4})}} \\ \vdots & \cdots & \cdots & \vdots \\ \frac{1}{1+e^{-(\mu_p+\tau_{p9})}} & \frac{1}{1+e^{-(\mu_p+\tau_{p9}+\eta_{p2})}} & \frac{1}{1+e^{-(\mu_p+\tau_{p9}+\eta_{p3})}} & \frac{1}{1+e^{-(\mu_p+\tau_{p9}+\eta_{p4})}} \end{bmatrix}.$$

If there is a one–to–one link, we require that

$$\begin{cases} \frac{1}{1+e^{-\mu_p}} & = & 0.0920980 \\ \frac{1}{1+e^{-(\mu_p+\eta_{p2})}} & = & 0.6557600 \\ \frac{1}{1+e^{-(\mu_p+\tau_{p3})}} & = & 0.0506880 \\ \frac{1}{1+e^{-(\mu_p+\tau_{p3}+\eta_{p2})}} & = & 0.6143500 \\ \vdots & & \\ \vdots & & \end{cases}.$$

The system has no solution. The first three equations can determine the values of μ_p, η_{p2} and τ_{p3}. But these estimates will not satisfy the fourth equation.

This is true for any model as long as there is one parameter with both time and heterogeneous effects. Using the linear link function, the differences between any two of the columns are constant and equal to $\eta_{\phi c}$ or η_{pc} if the model is heterogeneous in $\{\phi_{jc}\}$ or $\{p_{jc}\}$ respectively. However, when using the logistic link function, the differences between any two of the columns are no longer a constant. We have different point estimates from using the linear link function and using the logistic function. Therefore a different likelihood value results. □

6.4 Models With Interactions

We will now consider models with interactions between time and heterogeneous components. We will use the model $\{\phi(\cdot), p(t \times h_3)\}$ for illustration. For simplicity, we use the linear link function.

Example 6.4.1. This example shows that putting different constraints on the interaction matrix can give different point estimates. Therefore, a different likelihood value can be obtained.

The model $\{\phi(\cdot), p(t \times h_3)\}$ is constant in recapture probability and the survival probabilities are time–varying and heterogeneous. We also assume that there are interactions between time and heterogeneous components.

Using the linear link function, the point estimates of p_{jc} are given by

$$p_{jc} = \mu_p + \tau_{pj} + \eta_{pc} + (\tau\eta)_{pjc}, \tag{6.3}$$

with $\tau_{p2} = \eta_{p1} = 0$ and $p_{1c} = 1$ (see Notation 3.2.1). We first set $(\tau\eta)_{p2c} = (\tau\eta)_{pj1} = 0$, so that the interaction matrix is given by

$$[(\tau\eta)_{pjc}] = \begin{array}{c} \\ (\tau\eta)_{p1c} \\ (\tau\eta)_{p2c} \\ (\tau\eta)_{p3c} \\ (\tau\eta)_{p4c} \\ (\tau\eta)_{p5c} \\ (\tau\eta)_{p6c} \\ (\tau\eta)_{p7c} \\ (\tau\eta)_{p8c} \\ (\tau\eta)_{p9c} \end{array} \begin{array}{c} \begin{array}{ccc} (\tau\eta)_{pj1} & (\tau\eta)_{pj2} & (\tau\eta)_{pj3} \end{array} \\ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & (\tau\eta)_{p32} & (\tau\eta)_{p33} \\ 0 & (\tau\eta)_{p42} & (\tau\eta)_{p43} \\ 0 & (\tau\eta)_{p52} & (\tau\eta)_{p53} \\ 0 & (\tau\eta)_{p62} & (\tau\eta)_{p63} \\ 0 & (\tau\eta)_{p72} & (\tau\eta)_{p73} \\ 0 & (\tau\eta)_{p82} & (\tau\eta)_{p83} \\ 0 & (\tau\eta)_{p92} & (\tau\eta)_{p93} \end{bmatrix} \end{array}. \tag{6.4}$$

This determines the point estimates of p_{pj},

$$[p_{jc}] = \begin{bmatrix} 1 & 1 & 1 \\ \mu_p & \mu_p + \eta_{p2} & \mu_p + \eta_{p3} \\ \mu_p + \tau_{p3} & \mu_p + \tau_{p3} + \eta_{p2} + (\tau\eta)_{p32} & \mu_p + \tau_{p3} + \eta_{p3} + (\tau\eta)_{p33} \\ \mu_p + \tau_{p4} & \mu_p + \tau_{p4} + \eta_{p2} + (\tau\eta)_{p42} & \mu_p + \tau_{p4} + \eta_{p3} + (\tau\eta)_{p43} \\ \mu_p + \tau_{p5} & \mu_p + \tau_{p5} + \eta_{p2} + (\tau\eta)_{p52} & \mu_p + \tau_{p5} + \eta_{p3} + (\tau\eta)_{p53} \\ \mu_p + \tau_{p6} & \mu_p + \tau_{p6} + \eta_{p2} + (\tau\eta)_{p62} & \mu_p + \tau_{p6} + \eta_{p3} + (\tau\eta)_{p63} \\ \mu_p + \tau_{p7} & \mu_p + \tau_{p7} + \eta_{p2} + (\tau\eta)_{p72} & \mu_p + \tau_{p7} + \eta_{p3} + (\tau\eta)_{p73} \\ \mu_p + \tau_{p8} & \mu_p + \tau_{p8} + \eta_{p2} + (\tau\eta)_{p82} & \mu_p + \tau_{p8} + \eta_{p3} + (\tau\eta)_{p83} \\ \mu_p + \tau_{p9} & \mu_p + \tau_{p9} + \eta_{p2} + (\tau\eta)_{p92} & \mu_p + \tau_{p9} + \eta_{p3} + (\tau\eta)_{p93} \end{bmatrix}. \quad (6.5)$$

The maximum log–likelihood is found to be -836.515 with the point estimates of p_{jc} given by

$$[\hat{p}_{jc}] = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0.1968 & 0.8272 \\ 0.9523 & 0 & 1 \\ 0.9154 & 0.0158 & 0.2228 \\ 0.9431 & 0.0496 & 0.4206 \\ 0.8814 & 0 & 0.7441 \\ 0.8499 & 0.0374 & 0.2660 \\ 0.9039 & 0 & 0.4616 \\ 0.8861 & 0 & 0 \end{bmatrix}. \quad (6.6)$$

All values are rounded to 4 decimal places.

For comparison, we now set $(\tau\eta)_{p2c} = (\tau\eta)_{pj3} = 0$, so that the interaction

matrix is given by

$$[(\tau\eta)_{pjc}] = \begin{array}{c} \\ (\tau\eta)_{p1c} \\ (\tau\eta)_{p2c} \\ (\tau\eta)_{p3c} \\ (\tau\eta)_{p4c} \\ (\tau\eta)_{p5c} \\ (\tau\eta)_{p6c} \\ (\tau\eta)_{p7c} \\ (\tau\eta)_{p8c} \\ (\tau\eta)_{p9c} \end{array} \begin{array}{c} \begin{array}{ccc} (\tau\eta)_{pj1} & (\tau\eta)_{pj2} & (\tau\eta)_{pj3} \end{array} \\ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ (\tau\eta)_{p31} & (\tau\eta)_{p32} & 0 \\ (\tau\eta)_{p41} & (\tau\eta)_{p42} & 0 \\ (\tau\eta)_{p51} & (\tau\eta)_{p52} & 0 \\ (\tau\eta)_{p61} & (\tau\eta)_{p62} & 0 \\ (\tau\eta)_{p71} & (\tau\eta)_{p72} & 0 \\ (\tau\eta)_{p81} & (\tau\eta)_{p82} & 0 \\ (\tau\eta)_{p91} & (\tau\eta)_{p92} & 0 \end{bmatrix} \end{array} \tag{6.7}$$

with point estimates

$$[p_{jc}] = \begin{bmatrix} 1 & 1 & 1 \\ \mu_p & \mu_p + \eta_{p2} & \mu_p + \eta_{p3} \\ \mu_p + \tau_{p3} + (\tau\eta)_{p31} & \mu_p + \tau_{p3} + \eta_{p2} + (\tau\eta)_{p32} & \mu_p + \tau_{p3} + \eta_{p3} \\ \mu_p + \tau_{p4} + (\tau\eta)_{p41} & \mu_p + \tau_{p4} + \eta_{p2} + (\tau\eta)_{p42} & \mu_p + \tau_{p4} + \eta_{p3} \\ \mu_p + \tau_{p5} + (\tau\eta)_{p51} & \mu_p + \tau_{p5} + \eta_{p2} + (\tau\eta)_{p52} & \mu_p + \tau_{p5} + \eta_{p3} \\ \mu_p + \tau_{p6} + (\tau\eta)_{p61} & \mu_p + \tau_{p6} + \eta_{p2} + (\tau\eta)_{p62} & \mu_p + \tau_{p6} + \eta_{p3} \\ \mu_p + \tau_{p7} + (\tau\eta)_{p71} & \mu_p + \tau_{p7} + \eta_{p2} + (\tau\eta)_{p72} & \mu_p + \tau_{p7} + \eta_{p3} \\ \mu_p + \tau_{p8} + (\tau\eta)_{p81} & \mu_p + \tau_{p8} + \eta_{p2} + (\tau\eta)_{p82} & \mu_p + \tau_{p8} + \eta_{p3} \\ \mu_p + \tau_{p9} + (\tau\eta)_{p91} & \mu_p + \tau_{p9} + \eta_{p2} + (\tau\eta)_{p92} & \mu_p + \tau_{p9} + \eta_{p3} \end{bmatrix}. \tag{6.8}$$

We now obtain the maximum log–likelihood to be -832.999 with the point

estimates of p_{jc} given by

$$[\hat{p}_{jc}] = \begin{bmatrix} 1 & 1 & 1 \\ 0.1278 & 1 & 1 \\ 0 & 0.8547 & 1 \\ 0.0089 & 0.6860 & 0.9673 \\ 0.0380 & 0.7630 & 0.9767 \\ 0 & 0.6506 & 1 \\ 0.0419 & 0.5016 & 1 \\ 0 & 0.7640 & 0.9533 \\ 0 & 0.6734 & 0.9219 \end{bmatrix}. \tag{6.9}$$

All values are rounded to 4 decimal places.

In both cases, the maximum likelihood values are better than under the model $\{\phi(\cdot), p(t + h_3)\}$, which is -843.244. We see that by changing the constraints, we effectively have different point estimates, and therefore a different likelihood. □

Models with interactions can vary in many ways by changing the constraints. We will conclude by noting that they can improve the model fitting, but it is difficult to determine which constraints to use.

6.5 Conclusion

In this chapter, we have focused on fitting capture–recapture mixture models to data from open populations.

We have seen in Section 6.1 that optimization routines can sometimes stop at a local maximum if not enough initial points were taken to explore the parameter space. Without careful checking, we might miss the global optimum.

We also find that sometimes estimates that are close in the parameter space can give very different likelihood values. This is a challenge for the optimization routine and makes the fitting of the mixture models difficult.

We also discussed that for the model $\{[\phi(t+h), p(t+h)]_2\}$, where both the survival and the recapture probabilities are time–varying with two heterogeneous groups, it is not necessary to sub–divide the model into a "similar" and an "opposite" model by putting constraints on $\eta_{pc}\eta_{\phi c} > 0$ or $\eta_{pc}\eta_{\phi c} < 0$. Without such constraints, the optimization can identify the better model of the two.

We use the Akaike information criterion (AIC) to select the best models. By allowing up to four mixture components, we found better models for the Brushtail Possum (*Trichosurus vulpecula*) data, which were missed by the authors in PPN2003 as they have only fitted two mixture components.

We explored the effect of using different link functions in Section 6.3. Table 6.9 compares the AIC values from using the linear link function and using the logistic link function. We discussed a one–to–one transformation between the links. When there exists a one–to–one transformation, the results from using either link functions are equivalent, otherwise the results will be different.

We discussed models with interactions in Section 6.4. We illustrated using an example that by choosing different constraints on the interaction matrix, the point estimates can change, therefore resulting in a different likelihood. We saw here how the result from model fitting can change by having a different constraint.

To conclude, mixture models are difficult to fit to open population capture–recapture data. Careful checks are necessary, such as checking the gradients at the maximum likelihood estimates are zero (or close to 0). Profile likelihood

plots can also be used to check for global maximization. It is also essential to use multiple starting points.

6.6 List of files

In this section, we list all relevant files on the supplementary CD in Table 5.11. Details on using the Maple® package `MixtureModelforCaptureRecapture.mla` are given in Appendix A and details on using the Mathematica® package `MixtureModelforCaptureRecapture.m` are given in Appendix B.

All files are created and updated at the time of writing. PDF copies are given instead of the source files to avoid unintentionally editing the source.

File Name	Description
`hcjsRun.pdf`	File from the first author in PPN2003. Lines 69 – 91 show the number of iterations, cycles of fitting and the tolerances they have used.
`hcjsFunctions.pdf`	File from the first author in PPN2003. The file shows that the authors have only considered two mixture components.
`possfeb.pdf`	The fitted result from PPN2003.
`6.1.2_Combined_2nd.pdf`	Mathematica® notebook for Section 6.1.3 in PDF format.
`6.1.3_Female_S_0.pdf`	Mathematica® notebook for Section 6.1.4 in PDF format.
`6.1.4_Male_4C.pdf`	Mathematica® notebook for Section 6.1.5 in PDF format.
`CJSfit_Logistic_Link.pdf`	Model fitting using the logistic link function. Results for Table 6.6, Table 6.7 and Table 6.8.
`CJSfit_Linear_Link.pdf`	Model fitting using the linear link function. Results (of AIC values) are given in Table 6.9 for comparison.
`Logit_RandomSearch.m`	The maximum likelihood estimates from using the logistic link function and can be imported to Mathematica® .
`Linear_RandomSearch.m`	The maximum likelihood estimates from using the linear link function and can be imported to Mathematica® .
`6.4_Interactions.pdf`	Copy of Mathematica® notebook file for Section 6.4.

Table 6.10: List of files on Chapter 6

Chapter 7

Conclusion and Future Work

7.1 Conclusion

To conclude, in Chapter 2 we have extended the closed–population mixture models from Morgan and Ridout (2008). We presented the new mixture model of a binomial distribution with a logistic–normal binomial distribution. We found that when fitted to real data sets the performance of the new model is comparable to that of the binomial beta–binomial model in Morgan and Ridout (2008), and in some cases there were improvements.

In Chapter 3, we moved on to open population capture–recapture models. We looked in detail at the CJS capture–recapture mixture model from Pledger *et al.* (2003) and the JS capture–recapture mixture model from Pledger *et al.* (2010). We also reviewed the capture–recovery model and the capture–recapture–recovery model. We showed the format of the data, expressions with and without heterogeneity and examples of probabilities from corresponding capture histories for each of the models.

We started Chapter 4 by introducing the concept of *parameter redundancy* through a motivating example from the CJS model. We discussed

existing methods of detecting parameter redundancy using *exhaustive summaries*. We gave examples of using both the symbolic method and the hybrid symbolic–numerical (SN) method. We have shown examples of the parameter redundancy results from using the SN method for $K = 2, 3, \ldots 10$ and $C = 2, 3, \ldots 10$, using both the linear link function and the logistic link function. We compared our results with the closed population result from Holzmann *et al.* (2006), and we also compared our findings with these in Pledger *et al.* (2003). We noted the issue of obtaining different results from using different link functions. For finding general rules for parameter redundancy, we also listed the theorems that are need to develop the work further, and explained what we call *simpler exhaustive summaries*. A general framework for detecting parameter redundancy based on exhaustive summaries was adapted from Cole *et al.* (2010).

Chapter 5 focuses on exhaustive summaries and their use for determining the results of parameter redundancy. We started by looking at how to find a reparameterisation that is useful for developing the simpler exhaustive summary using the linear link function. We then provided the fine detail for various examples, from the relatively simple ones to more complicated models. We also gave examples for models using the logistic link function and explained the relevance of one–to–one transformations. After that we discussed how to use the simpler exhaustive summary to obtain general conditions for parameter redundancy and listed tables of parameter redundancy results.

We returned to model fitting in Chapter 6. We fitted a range of plausible models, except the models with interactions, to the Australian brushtail possum (*Trichosurus vulpecula*) data, kindly provided by Professor Pledger. The results were compared with those from Pledger *et al.* (2003). We note that we have obtained better results. Not having enough starting points for

maximum likelihood optimization could result in a local optimum and stop the optimization from finding a global maximum to the likelihood. We also noted that by allowing up to four mixture components, we have found structurally simpler models than those found in Pledger *et al.* (2003) for each of the combined data, female data and male datasets. We demonstrated that different results can be obtained by using different link functions if there is no one–to–one transformation between the links. We showed that the results can also be affected by changing the constraints on the interaction matrix for models with interactions between time and heterogeneous components.

7.2 Future Work

We have seen that it is not easy to know whether we can estimate all parameters of a mixture model. Even with the existing formal methods, we rely heavily on the use of symbolic computer packages. There is still work to be done when the logistic link function is being used as well as looking at models with interactions. There are also a lot to explore on the fitting of such mixture models, specifically the effect of having different valid constraints. To finish off the thesis, we again quote the following from the web supplementary material in Pledger *et al.* (2010):

> Finding exact rules for these more complex open–population models would be a difficult and time–consuming exercise, and the rules are unlikely to be simple.

This sentence best describes the majority of the work in this thesis and the nature of the work, which is in general not easy to conduct.

Appendix A

Maple® Computer Package: `MixtureModelforCaptureRecapture`

A.1 About the package

The Maple® package `MixtureModelforCaptureRecapture.mla` is written specifically to check the parameter redundancy results for the models described in Chapter 3.

A.2 How to use the package?

If you have Maple® installed on your computer, the compiled package `MixtureModelforCaptureRecapture.mla` can be added on by double clicking. After this, type `with(MixtureModelforCaptureRecapture)` to load the package into the Maple® session, see Figure A.1.

The `libname:=` reviews your local path to the Maple® library and the path where this package is currently being loaded from. If you want to permanently install the package to your copy of Maple® , this can be done by copying the package (.mla) file to your Maple® library folder. If you want to delete the package from your Maple® library, this is done by deleting (or removing) the


```
> march('open',
  "D:\\Dropbox\\PhD_thesis\\Codes\\chapter5\\Mix
  tureModelforCaptureRecapture.mla");
libname := "C:\Program Files\Maple 18\lib", ".",                    (1)
    "D:\Dropbox\PhD_thesis\Codes\chapter5\MixtureModelforCa\
    ptureRecapture.mla"
> with(MixtureModelforCaptureRecapture):
> myMixtureModelforCaptureRecapture();
Welcome to Maple package
'MixtureModelforCaptureRecapture'!
This is version 1.0.
Last updated on 07 March 2015.
```

Figure A.1: How to install the Maple® package?

package (.mla) file from your Maple® library path.

A.3 Functions Available

We give a short summary of the functions available in Table A.1. Some functions are created solely for shorthand, avoiding typing in long syntax repeatedly.

The source code of all procedures can be obtained by using

```
1    print(funcname);
2    # or
3    showstat(funcname);
```

where `funcname` is the name of the procedure. For example,

```
1  > print(myCHcc);
2  proc (K::integer, { M::truefalse := false }, ‘ $‘)
3      local CH::list;
4      description " Capture History for Capture Recapture ";
5      CH := combinat[permute]([seq(‘$‘(s, K), s = [0, 1])], K);
6      if M then return Matrix(sort(CH)) else return sort(CH) end if
7  end proc
```

This prints the source of the procedure `myCHcc`, which gives the full capture histories of a K–year capture–recapture study.

Maple Syntax	Description
myCHcc	full history for capture–recapture
myCHcr	full history for capture–recovery
myCHcrr	full history for capture–recapture–recovery
myparM	gives the matrices (in order) of ϕ_{jc}, p_{jc} and λ_{jc}
myinteractM	gives the interaction terms
myindprobcc	gives the probability from a capture–recapture history
myindprobcr	gives the probability from a capture–recovery history
myindprobcrr	gives the probability from a capture–recapture–recovery history
mykappa	gives the original exhaustive summary
mypars	gives all parameters in the exhaustive summary
myDD1	gives the derivative matrix
myEstpars	gives estimable parameters (for parameter redundant models only)
mysscc	reparametrisation for capture–recapture models
sskappa	simplifying kappa using ss

Table A.1: Functions in the Maple® package MixtureModelforCaptureRecapture.mla

A.4 The mykappa function

The main function to generate an exhaustive summary is mykappa. For other functions, detailed explanations are given in the Maple® worksheet when they first appear. We will focus on the use of mykappa in this section.

To see what inputs mykappa takes, we use

```
> showstat(mykappa);
```

and the first few lines from the output are given below:

```
MixtureModelforCaptureRecapture:-mykappa := proc(
    K::integer, C::integer, CH::list, {
        CC::truefalse := true,
        CR::truefalse := false,
        CRR::truefalse := false,
        JS::truefalse := false,
        hlambda::truefalse := false,
        hp::truefalse := false,
        hphi::truefalse := false,
        intlambda::truefalse := false,
        intp::truefalse := false,
        intphi::truefalse := false,
        logit::truefalse := false,
        logs::truefalse := false,
        printparM::truefalse := false,
        tvarlambda::truefalse := false,
        tvarp::truefalse := false,
        tvarphi::truefalse := false
}, $)

# with the rest omitted
```

The first two inputs K and C are integer values as given in the text. K is the length of the open population capture–recapture study and C is the number of mixture components. The third input `CH` is the capture history matrix. By default, it generates a complete capture history matrix from the model you have input (which is the CJS mixture by default).

The arguments between the curly brackets {} are optional, all with a value set to either `true` or `false`. If the values are omitted, by default they have a value `false` except `CC`. Table A.2 gives the explanation for each options.

A.5 Using a real dataset

The function `mykappa` also accepts a capture–history matrix `CH`. For example, instead of using the full capture history for $K = 3$, we can define a list of capture histories as the following:

```
    mykappa(3,1,tvarp,tvarphi); # CJS with K=3 (full history)

    # Define my own history (as a list)
    CH:=[[1,1,1],[1,1,1],[1,1,1],[0,1,0]];
    # Finding the probabilities
    mykappa(3,1,CH,tvarp,tvarphi);
```

Maple® Syntax	Description
CC	CC=true for capture–recapture model by default
CR	CR=true for capture–recovery model
CRR	CRR=true for capture–recapture–recovery model
JS	JS=true for Jolly–Seber mixture models
hlambda	hlambda=true when the recovery probabilities are heterogeneous
hp	hp=true when the recapture probabilities are heterogeneous
hphi	hphi=true when the survival probabilities are heterogeneous
intlambda	intlambda=true when there are interaction terms in the recovery probabilities
intp	intp=true when there are interaction terms in the recovery probabilities
intphi	intphi=true when there are interaction terms in the recovery probabilities
logit	logit=true when the logistic link function is used (by default, the linear link function is used)
logs	logs=true when the log link function is used
printparM	printparM=true show the parameter matrices for $\phi_{jc}, p_{jc}, \lambda_{jc}$
tvarlambda	tvarlambda=true when the recovery probabilities are time–varying
tvarp	tvarp=true when the recapture probabilities are time–varying
tvarphi	tvarphi=true when the survival probabilities are time–varying

Table A.2: Optional arguments for Maple® function mykappa

.

By omitting the `CH`, we assume a full capture history. That is, line 1 is identical to having the following

```
1    mykappa(3,1,myCHcc(3),tvarp,tvarphi);
```

Note since by default we also assume it is a CJS mixture model, the first capture history `[0,0,0]` from `myCHcc(3)` is ignored.

Appendix B

Mathematica® Computer Package:

MixtureModelforCaptureRecapture

B.1 About the package

The Mathematica® package `MixtureModelforCaptureRecapture.m` is written specifically to check the parameter redundancy results for models described in Chapter 3.

B.2 How to use the package?

If you have Mathematica® installed on your computer, the compiled package `MixtureModelforCaptureRecapture.m` can be imported by entering the following:

```
1    SetDirectory[NotebookDirectory[]];
2    Get["MixtureModelforCaptureRecapture.m", "ChenYu_thesis"]
```

if you have *saved* your notebook and the package in the same folder.

B.3 Functions Available

We give a short summary of the functions available in Table B.1. Some functions are created solely for shorthand, avoiding typing in long syntax repeatedly.

The source code of all procedures can be obtained by using

```
?funcname
(* or *)
??funcname
```

where `funcname` is the name of the procedure. For example,

```
??myCH

    Gives full capture history from a K--year study.
    myCH[CJS][K_Integer/;K>=0]:=Rest[Tuples[{0,1},K]]

    myCH[JS][K_Integer/;K>=0]:=Tuples[{0,1},K]

    myCH[CR][K_Integer/;K>=0]:=Sort[
    ReplaceList[ConstantArray[0,K],{
    Private`a___,Private`x_,
    Private`b___,Private`y_,
    Private`c___
    }:>{Private`a,1,Private`b,2,Private`c}]]

    myCH[CRR][K_Integer/;K>=0]:=Sort[Join@@Table[If[
    Private`k==0,Identity,Join[Join[#1,{2}],
    ConstantArray[0,Private`k-1]]&]/@Rest[
    Tuples[{0,1},K-Private`k]
    ],{Private`k,0,K-1}]]
```

This prints the source of the procedure `myCH`, which gives the full capture histories of a K–year capture–recapture study.

B.4 Usage

The usage of the Mathematica® package is similar to the Maple® package described in Appendix A. The main function is `mykappa`, similar to the Maple® function described in Section A.4. Options of the Mathematica® function are given in Table B.2.

The function `mykappa` takes three compulsory inputs in the order of K, C and CH, where K is the length of study (as defined in text), C is the number

Maple Syntax	Description
myCH["CJS"][K]	gives full capture histories from a K–year CJS capture–recapture study;
myCH["JS"][K]	gives full capture histories from a K–year JS capture–recapture study;
myCH["CR"][K]	gives full capture histories from a K–year capture–recovery study;
myCH["CRR"][K]	gives full capture histories from a K–year capture–recapture–recovery study;
myparM	gives the matrices (in order) of ϕ_{jc}, p_{jc} and λ_{jc}
myinteractM	gives the interaction terms
myprobCJS	gives CJS mixture probability
myprobJS	gives JS mixture probability
myprobCR	gives capture–recovery mixture probability
myprobCRR	gives capture–recapture–recovery mixture probability
mykappa	gives the original exhaustive summary
mynum	gives the model rank using the SN method
myDD1	gives the derivative matrix
myEstpars	gives estimable parameters (for parameter redundant models only)
H	the H function
makess	gives the reparameterisation
sskappa	simplify using the reparameterisation
myoptlogit	Optimization routine for using the logistic link function
myoptlinear	Optimization routine for using the linear link function

Table B.1: Functions in the Mathematica® package MixtureModelforCaptureRecapture.m

Mathematica® Syntax	Description
hlambda	hlambda -> true when the recovery probabilities are heterogeneous
hp	hp -> true when the recapture probabilities are heterogeneous
hphi	hphi -> true when the survival probabilities are heterogeneous
intlambda	intlambda -> true when there are interaction terms in the recovery probabilities
intp	intp -> true when there are interaction terms in the recovery probabilities
intphi	intphi -> true when there are interaction terms in the recovery probabilities
logit	logit -> true when the logistic link function is used (by default, the linear link function is used)
tvarlambda	tvarlambda -> true when the recovery probabilities are time–varying
tvarp	tvarp -> true when the recapture probabilities are time–varying
tvarphi	tvarphi -> true when the survival probabilities are time–varying

Table B.2: Optional arguments for Mathematica® function mykappa
.

of mixture components and CH is a capture history matrix. It determines the model by the capture histories you have entered.

For example, using the combined Brushtail Possum (*Trichosurus vulpecula*) data from Section 6.1, we enter

```
(* Loading the package *)
SetDirectory[NotebookDirectory[]];
Get["MixtureModelforCaptureRecapture.m", "ChenYu_thesis"]
(* Import data *)
data = Import["combined_unique_sort.csv", "CSV"];
CH = data[[All, 1 ;; 9]];
feq = data[[All, 10]];
```

The data is stored as a comma separated values (.csv) file, where the first 9 columns are entries of 0s and 1s and the last column (column 10) are the frequencies of each such row history.

We can now get the CJS mixture probabilities by entering

```
mykappa[9, 2, CH, hp -> True]
```

for the model $\{\phi(\cdot), p(h_2)\}$ for example. By default, the logistic link function is used. For more usage, please refer to Mathematica® notebooks in the supplementary materials.

References

Abramowitz, M. and Stegun, I. (1964) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover.

Aitchison, J. and Shen, S. M. (1980) Logistic–Normal distributions: Some properties and uses. *Biometrika*, **67**, 261–272.

Akaike, H. (1974) A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, **19**, 716–723.

Amstrup, S. C., McDonald, T. L. and Manly, B. F. J. (2005) *Handbook of capture–recapture Analysis*. Princeton University Press.

Anderson, D. G. (1965) Gaussian quadrature formulae for $\int_0^1 -\ln(x) f(x)\,\mathrm{d}x$. *Mathematics of Computation*, **19**, 477–481.

Baillargeon, S. and Rivest, L.-P. (2007) Rcapture: Loglinear models for capture–recapture in R. *Journal of Statistical Software*, **19**.

Borchers, D. L., Stephens, S. T. and Zucchini, W. (2002) *Estimating Animal Abundance: Closed Populations*. Springer London.

Boulinier, T., Nichols, J. D., Sauer, J. R., Hines, J. E. and Pollock, K. H. (1998) Estimating species richness: The importance of heterogeneity in species detectability. *Ecology*, **79**, 1018–1028.

Burnham, K. P. (1972) *Estimation of population size in multiple capture re-*

capture studies when capture probabilities vary among animals. Ph.D. thesis, Oregon State University.

Byrd, R. H., Lu, P., Nocedal, J. and Zhu, C. (1995) A limited memory algorithm for bound constrained optimization. *SIAM J. Scientific Computing*, **16**, 1190–1208.

Carothers, A. D. (1973) Capture–recapture methods applied to a population with known parameters. *Journal of Animal Ecology*, **42**, 125–146.

Caswell, H., Hunter, C. M., Thomson, D. L., Cooch, E. G. and Conroy, M. J. (2009) *Modeling Demographic Processes In Marked Populations*, chap. Rank and Redundancy of Multistate Mark–Recapture Models for Seabird Populations with Unobservable States, 797–825. Springer.

Catchpole, E. A. and Morgan, B. J. T. (1997) Detecting parameter redundancy. *Biometrika*, **84**, 187–196.

Catchpole, E. A. and Morgan, B. J. T. (2001) Deficiency of parameter redundant models. *Biometrika*, **88**, 593–598.

Catchpole, E. A., Morgan, B. J. T. and Freeman, S. N. (1998) Estimation in parameter–redundant models. *Biometrika*, **85**, 462–468.

Catchpole, E. A., Morgan, B. J. T. and Viallefont, A. (2002) Solving problems in parameter redundancy using computer algebra. *Journal of Applied Statistics*, **29**, 625–636.

Chapman, D. G. (1951) *Some properties of the hypergeometric distribution with applications to zoological sample censuses*. University of California Press.

Choquet, R. and Cole, D. J. (2012) A hybrid symbolic–numerical method for determining model structure. *Mathematical Biosciences*, **236**, 117–125.

Choquet, R. and Gimenez, O. (2012) Towards built–in capture–recapture mixed models in program E–SURGE. *Journal of Ornithology*, **152**, 625–639.

Cole, D. J. (2012) Determining parameter redundancy of multi–state mark–recapture models for sea birds. *Journal of Ornithology*, **152**, 305–315.

Cole, D. J. and Morgan, B. J. T. (2010a) A note on determining parameter redundancy in age–dependent tag return models for estimating fishing mortality, natural mortality and selectivity. *JABES*, **15**, 431–434.

Cole, D. J. and Morgan, B. J. T. (2010b) Parameter redundancy with covariates. *Biometrika*, **97**, 1002–1005.

Cole, D. J., Morgan, B. J. T., Catchpole, E. A. and Hubbard, B. A. (2012) Parameter redundancy in mark–recovery models. *Biometrikal Journal*, **54**, 507–523.

Cole, D. J., Morgan, B. J. T., Choquet, R., McCrea, R. S., Pradel, R. and Gimenez, O. (2014) Does your species have memory? Analyzing capture–recapture data with memory models. *Ecology and Evolution*, **4**, 2124–2133.

Cole, D. J., Morgan, B. J. T. and Titterington, D. M. (2010) Determining the parametric structure of models. *Mathematical Biosciences*, **228**, 16–30.

Corless, R. M. and Jeffrey, D. J. (1997) The Turing factorization of a rectangular matrix. *SIGSAM Bulletin*, **31**, 20–30.

Cormack, R. M. (1964) Estimates of survival from the sighting of marked animals. *Biometrika*, **51**, 429–438.

Coull, B. A. and Agresti, A. (1999) The use of mixed logit models to reflect heterogeneity in capture–recapture studies. *Biometrics*, **55**, 294–301.

Crosbie, S. F. and Manly, B. F. J. (1985) Parsimonious modelling of capture–mark–recapture studies. *Biometrics*, **41**, 385–398.

Crouch, E. A. C. and Spiegelman, D. (1990) The evaluation of integrals of the form $\int_{-\infty}^{\infty} f(t)\exp(-t^2)\,\mathrm{d}t$: Application to logistic–normal models. *Journal of the American Statistical Association*, **85**, 464–469.

Darroch, J. N. (1958) The Multiple–Recapture Census: I. Estimation of a Closed Population. *Biometrika*, **45**, 343–359.

Darroch, J. N. (1959) The Multiple–Recapture Census: II. Estimation When There is Immigration or Death. *Biometrika*, **46**, 336–351.

Davis, P. J. and Rabinowitz, P. (1984) *Methods of Numerical Integration*. Academic Press.

Dorazio, R. M. and Royle, J. A. (2003) Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics*, **59**, 351–364.

Edwards, W. R. and Eberhardt, L. (1967) Estimating cottontail abundance from livetrapping data. *The Journal of Wildlife Management*, **31**, 87–96.

Gimenez, O. and Choquet, R. (2010) Individual heterogeneity in studies on marked animals using numerical integration: capture–recapture mixed models. *Ecology*, **91**, 951–957.

Gimenez, O., Choquet, R. and Lebreton, J. D. (2003) Parameter redundancy in multistate capture–recapture models. *Biometrikal Journal*, **45**, 704–722.

Gimenez, O., Viallefont, A., Catchpole, E. A., Choquet, R. and Morgan, B. J. (2004) Methods for investigating parameter redundancy. *Animal Biodiversity and Conservation*, **27**, 561–572.

González, J., Tuerlinckx, F., Boeck, P. D. and Cools, R. (2006) Numerical integration in logistic–normal models. *Journal Computational Statistics & Data Analysis*, **51**, 1535–1548.

Greenwood, R. E. and Miller, J. J. (1984) Zeros of the Hermite polynomials and weights for Gauss' mechanical quadrature formula. *Bulletin of the American Mathematical Society*, **54**, 765–769.

Holzmann, H., Munk, A. and Zucchini, W. (2006) On identifiability in capture–recapture models. *Biometrics*, **62**, 934–936.

Hubbard, B. A. (2014) *Parameter Redundancy with Applications in Statistical Ecology*. Ph.D. thesis, University of Kent.

Hubbard, B. A., Cole, D. J. and Morgan, B. J. T. (2014) Parameter redundancy in capture–recapture–recovery models. *Statistical Methodology*, **17**, 17–29.

Hunter, C. M., Caswell, H., Thomson, D. L., Cooch, E. G. and Conroy, M. J. (2009) *Rank and Redundancy of Multistate Mark–Recapture Models for Seabird Populations with Unobservable States*, chap. IX, 797–825. Springer US.

Jiang, H. H., Pollock, K. H., Brownie, C., Hoenig, J. M., Latour, R. J., Wells, B. K. and Hightower, J. E. (2007) Tag return models allowing for harvest and catch and release: Evidence of environmental and management impacts on striped bass fishing and natural mortality rates. *North American Journal of Fisheries Management*, **27**, 387–396.

Jolly, G. M. (1965) Explicit estimates from capture–recapture data with both death and immigration–stochastic model. *Biometrika*, **52**, 225–247.

King, R. and Brooks, S. (2008) On the Bayesian estimation of a closed population size in the presence of heterogeneity and model uncertainty. *Biometrics*, **64**, 816–824.

King, R., Brooks, S. and Coulson, T. (2008) Analyzing complex capture–recapture data in the presence of individual and temporal covariates and model uncertainty. *Biometrics*, **64**, 1187–1195.

King, R., Morgan, B. J. T., Gimenez, O. and Brooks, S. (2009) *Bayesian Analysis for Population Ecology*. Chapman and Hall, London.

Kirkpatrick, S., Gelatt, C. D. and Vecchi, M. P. (1983) Optimization by simulated annealing. *Science*, **220**, 671–680.

Lebreton, J.-D., Burnham, K. P., Clobert, J. and Anderson, D. R. (1992) Modeling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecological Monographs*, **62**, 67–118.

Lincoln, F. C. (1930) Calculating waterfowl abundance on the basis of banding returns. *United States Department of Agriculture Circular*, **118**, 1 – 4.

Link, W. A. (2003) Nonidentifiability of population size from capture–recapture data with heterogeneous detection probabilities. *Biometrics*, **59**, 1123–1130.

Liu, Q. and Pierce, D. A. (1994) A note on Gauss–Hermite quadrature. *Biometrika*, **81**, 624–629.

Manly, B. F. J. and Parr, M. J. (1968) A new method of estimating population size, survivorship, and birth rate from capture–recapture data. *Transactions of the Society for British Entomology*, **18**, 81–89.

Mares, M. A., Streilein, K. E. and Willig, M. R. (1981) Experimental assessment of several population estimation techniques on an introduced population of eastern chipmunks. *Journal of Mammalogy*, **62**, 315–328.

Mathar, R. J. (2013) Gauss–Laguerre and Gauss–Hermite Quadrature on 64, 96 and 128 Nodes. *Functions and Analysis*.

McCrea, R. S. and Morgan, B. J. T. (2011) Multistate mark–recapture model selection using score tests. *Biometrics*, **67**, 234–241.

McCrea, R. S. and Morgan, B. J. T. (2014) *Analysis of capture–recapture Data*. Chapman and Hall, Boca Raton.

McCrea, R. S., Morgan, B. J. T. and Cole, D. J. (2013) Age–dependent mixture models for recovery data on animals marked at unknown age. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **62**, 101–113.

Moiseev, S. (2011) Universal derivative–free optimization method with quadratic convergence. *Optimization and Control*.

Morgan, B. J. T. (2008) *Applied Stochastic Modelling*. Chapman and Hall/CRC.

Morgan, B. J. T. and Ridout, M. S. (2008) A new mixture model for capture heterogeneity. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **57**, 433–446.

Morgan, B. J. T. and Ridout, M. S. (2009) Estimating N: a robust approach to capture heterogeneity. In *Modeling Demographic Processes in Marked Populations*, vol. 57, 1069–1080. Springer.

Myint-U, T. and Debnath, L. (2007) *Linear Partial Differential Equations for Scientists and Engineers*. Birkhauser Boston.

Nelder, J. A. and Mead, R. (1965) A simplex algorithm for function minimization. *Computer Journal*, **7**, 308–313.

Nichols, J. D., Pollock, K. H. and Hines, J. E. (1984) The use of a robust capture–recapture design in small mammal population studies: A field example with microtus pennsylvanicus. *Acta Theriologica*, **29**, 357–365.

Nixon, C. M., Edwards, W. R. and Eberhardt, L. (1967) Estimating squirrel abundance from livetrapping data. *The Journal of Wildlife Management*, **31**, 96–101.

Otis, D. L., Burnham, K. P., White, G. C. and Anderson, D. R. (1978) Statistical inference from capture data on closed animal populations. *Wildlife Monographs*, **62**, 3–135.

Pawitan, Y. (2001) *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford University Press, Oxford.

Petersen, C. G. J. (1896) The yearly immigration of young plaice into the limfjord from the German sea. *Report of the Danish Biological Station*, **6**, 5–84.

Phillpot, P. (2000) *The Skinks of North Brother Island: Abundance, Habitat Use and Species Interactions*. Ph.D. thesis, Victoria University of Wellington, NZ.

Pledger, S. (2000) Unified maximum likelihood estimates for closed capture–recapture models using mixtures. *Biometrics*, **56**, 434–442.

Pledger, S. (2005) The performance of mixture models in heterogeneous closed population capture–recapture. *Biometrics*, **61**, 868–873.

Pledger, S. and Phillpot, P. (2008) Using mixtures to model heterogeneity in ecological capture–recapture studies. *Biometrikal Journal*, **50**, 1022–1034.

Pledger, S., Pollock, K. H. and Norris, J. L. (2003) Open capture–recapture models with heterogeneity: I. Cormack–Jolly–Seber model. *Biometrics*, **59**, 786–794.

Pledger, S., Pollock, K. H. and Norris, J. L. (2010) Open capture–recapture models with heterogeneity: II. Jolly–Seber model. *Biometrics*, **66**, 883–890.

Pledger, S. and Schwarz, C. J. (2002) Modelling heterogeneity of survival in band–recovery data using mixtures. *Journal of Applied Statistics*, **29**, 315–327.

Pollock, K. H., Nichols, J. D., Brownie, C. and Hines, J. E. (1990) Statistical inference for capture–recapture experiments. In *Wildlife Monographs*.

Read, C. B., Balakrishnan, N., Vidakovic, B. and Kotz, S. (2006) *Encyclopaedia of Statistical Sciences*. Wiley–Blackwell.

Ridout, M. S. (2007) Software for fitting capture heterogeneity models to estimate the size of a closed population. Tech. rep., University of Kent, Canterbury.

Robbins, C., Bystrak, D. and Geissler, P. (1986) *The Breeding Bird Survey: Its First Fifteen Years, 1965–1979*. Ph.D. thesis, Federal Government Series: Resource Publication.

Rothenberg, T. J. (1971) Identification in parametric models. *Econometrica*, 577–591.

Royston, P. (2007) Profile likelihood for estimation and confidence intervals. *The Stata Journal*, **7**, 376–387.

Salzer, H., Zucker, R. and Capuano, R. (1952) Table of the zeros and weight factors of the first 20 Hermite polynomials. *Journal of Research of the National Institute of Standards and Technology*, **48**, 111–116.

Salzer, H. E. and Zucker, R. (1949) Table of the zeros and weight factors of the first fifteen Laguerre polynomials. *Bulletin of the American Mathematical Society*, **55**, 1004–1012.

Schnabel, Z. E. (1938) The estimation of total fish population of a lake. *The American Mathematical Monthly*, **45**, 348–352.

Schwarz, C. J. and Arnason, A. N. (1996) A general methodology for the analysis of capture–recapture experiments in open populations. *Biometrics*, **52**, 860–873.

Seber, G. A. F. (1965) A note on the multiple recapture census. *Biometrika*, **52**, 249–259.

Shao, T. S., Chen, T. C. and Frank, R. M. (1964) Tables of zeros and Gaussian weights of certain associated Laguerre polynomials and the related generalized Hermite polynomials. *Mathematics of Computation*, **18**, 598–616.

Silvey, S. (1975) *Statistical Inference*. CRC press.

Stafford, J. E. and Andrews, D. F. (1993) A symbolic algorithm for studying adjustments to the profile likelihood. *Biometrika*, **80**, 715–730.

Storn, R. and Price, K. (1995) Differential evolution: A simple and efficient adaptive scheme for global optimization over continuous spaces. *Journal of Global Optimization*, **11**, 341–359.

Stroud, A. H. and Secrest, D. (1966) *Gaussian Quadrature Formulas*. Prentice Hall.

Sutherland, W. J. (1996) *Ecological Census Techniques: A Handbook*. Cambridge University Press, Cambridge.

Tanton, M. T. (1965) Problems of live–trapping and population estimation for the wood mouse, Apodemus sylvaticus (L.). *Journal of Animal Ecology*, **34**, 1–22.

Venzon, D. J. and Moolgavkar, S. H. (1988) A method for computing profile–likelihood–based confidence intervals. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **37**, 87–94.

Viallefont, A., Lebreton, J. D., Reboulet, A.-M. and Gory, G. (1998) Parameter identifiability and model selection in capture–recapture models: A numerical approach. *Biometrikal Journal*, **40**, 313–325.

White, G. C., Anderson, D. R. Burnham, K. P. and Otis, D. L. (1982) Capture–recapture and removal methods for sampling closed populations. Tech. Rep. Report LA–8787–NERP, Los Alamos National Laboratory. 235pp.

Whittaker, E. T. and Watson, G. N. (1990) *A Course of Modern Analysis*. Cambridge University Press, Cambridge.

Zhigljavsky, A. A. (1991) *Theory of Global Random Search*. Springer.

[This page is intentionally left blank.]